UNIVERSITEIT·STELLENBOSCH·UNIVERSITY
jou kennisvennoot • your knowledge partner

SET THEORY

P. Ouwehand

Department of Mathematical Sciences
Stellenbosch University

2012

ZFC

ZFC 0. Set Existence: There is a set.

$$\exists x(x = x)$$

ZFC 1. Extensionality: If sets x, y have the same elements, then $x = y$.

$$\forall x\ \forall y\ [\forall z(z \in x \leftrightarrow z \in y) \rightarrow x = y]$$

ZFC 2.$_{\varphi}$ **Separation Schema:** If $P(x)$ is a property of sets (with parameter p) describable by a first–order sentence $\varphi(x, p)$, then for any set Z and any p there is a set $Y = \{x \in Z : P(x)\}$ which consists precisely of those elements of Z which have the property P.

$$\forall z\ \forall p\ \exists y\ \forall x\ (x \in y \leftrightarrow x \in z \wedge \varphi(x, p))$$

ZFC 3. Pairing: For any x, y there is a set $z = \{x, y\}$ whose elements are precisely x and y.

$$\forall x\ \forall y\ \exists z\ [\forall w\ (w \in z \leftrightarrow w = x \vee w = y)]$$

ZFC 4. Union: For every set x there is a set $y = \bigcup x$ which is the union of all the elements of x.

$$\forall x\ \exists y\ \forall z\ [z \in y \leftrightarrow \exists w \in x\ (z \in w)]$$

ZFC 5. Power Set: For any set x there is a set $y = \mathcal{P}(x)$ whose elements are precisely all the subsets of x.

$$\forall x\ \exists y\ \forall z\ (z \in y \leftrightarrow z \subseteq x)$$

ZFC 6. Infinity: There is an infinite set.

$$\exists x\ [\varnothing \in x \wedge \forall y\ (y \in x \rightarrow y \cup \{y\} \in x)]$$

ZFC 7. Choice: Every family of non–empty sets has a *choice function*: If x is a family of non–empty sets, then there is a function f on x which *chooses* from each $w \in x$ an element $f(w) \in w$.

$$\forall x\ [\forall w \in x\ (w \neq \varnothing) \rightarrow \exists f\ (\ f \text{ is a function} \wedge \operatorname{dom}(f) = x \wedge \forall w \in x\ (f(w) \in w))]$$

ZFC 8.$_{\varphi}$ **Replacement Schema:** A relation $R(u, v)$ between sets which is describable by a first–order sentence $\varphi(u, v, p)$ (with parameter p) is said to be *functionlike* if for every u there is at most one v such that $R(u, v)$.
If R is functionlike and x is a set, then the image $z = R[x]$ of x under R is a set.

$$\forall p\ \Big[\Big(\forall u\ \forall v\ \forall w\ [\varphi(u, v, p) \wedge \varphi(u, w, p) \rightarrow v = w]\Big) \rightarrow \forall x\ \exists y\ \forall v\ (v \in y \leftrightarrow \exists u \in x\ \varphi(u, v, p))\Big]$$

ZFC 9. Foundation: Every non–empty set has a $\in$–minimal element.

$$\forall x\ [x \neq \varnothing \rightarrow \exists y \in x\ (y \cap x) = \varnothing]$$

These notes are for a short course in set theory at the undergraduate level at Stellenbosch University. No pretense at orignality is claimed. Though amplified by material from a number of additional sources, the debt to the first few chapters of the book *Set Theory*, by Thomas Jech, Springer 2003, should be easily discernible.

Contents

Chapter 1

The Axioms of Set Theory

1.1 Introduction

Every mathematician needs to know some set theory. “Set theory”, according to (the set theorist) Kenneth Kunen, “is a theory of everything”. *All* of mathematics can be done within the framework of set theory. This means that all mathematical objects can be *interpreted* as sets, and all mathematical statements about these objects can be phrased in the language of sets. Not only *can* this be done: Typically, it *is* done.

We assume here that, to some extent, you are already aware of this. For example, a group is an ordered tuple $(G, \cdot, ^{-1}, e)$, where

- G is a set.
- $\cdot : G \times G \to G$ is a function which plays the role of group multiplication.
- $^{-1} : G \to G$ is a function which plays the role of multiplicative inverse
- $e \in G$ is an element which plays the role of multiplicative identity element.

You probably know that a function $f : A \to B$ is a set: It a subset of the Cartesian product $A \times B$, where we regard $(a, b) \in f$ if and only if $f(a) = b$. The Cartesian product $A \times B$ of two sets is itself a set, namely the set of ordered pairs (a, b) where $a \in A$ and $b \in B$. An ordered pair (a, b) can be interpreted as a set also, namely $(a, b) := \{\{a\}, \{a, b\}\}$. Ordered tuples of higher order can be defined iteratively,

$$(a, b, c) := ((a, b), c) = \{\{(a, b)\}, \{(a, b), c\}\} = \{\{\{\{a\}, \{a, b\}\}\}, \{\{\{a\}, \{a, b\}\}, c\}\}$$

The group laws can then be written as equality of certain functions: For example, the associative law $(ab)c = a(bc)$ is the assertion that the functions given by the compositions

$$G \times G \times G \to G \times G \to G : (a, b, c) \mapsto (ab, c) \mapsto (ab)c$$

and

$$G \times G \times G \mapsto G \times G \mapsto G : (a, b, c) \to (a, bc) \to a(bc)$$

are equal. Since functions are sets, the associative law for G is equivalent to the assertion that two sets are equal. In principle, everything about groups can be formulated in the language of sets. It will look horrendous, it is true, but it can be done. Thus the whole group $(G, \cdot, ^{-1}, e)$

is a set — not just the base set G on which the group operations are defined — the whole object $(G, \cdot, ^{-1}, e)$. And each element of $(G, \cdot, ^{-1}, e)$ is a set,. . . and each element of an element of $(G, \cdot, ^{-1}, e)$ is a set, and. . . ,

In the same way all the objects you encounter in mathematics — natural numbers, complex numbers, topological spaces, differentiable manifolds, tangent bundles, Boolean algebras, measure spaces, stochastic processes, pseudo–differential operators — can be *interpreted* as sets. We do not say that these objects *are* sets; indeed, we prefer to completely avoid metaphysical speculation about what mathematical objects might *be*, and in what sense they might be said to *exist*. We simply mean that we can construct sets that *behave* like these mathematical objects. This kind of thing is not new to you: You're probably quite used to regarding real numbers as points on a line. But a point is not the same thing as a real number. "A point," says Euclid, "is that which has no parts". A real number, on the other hand, is an element of a complete ordered field. But, thought of in the right way, the points on a line can be said to behave like the real numbers, and thus can be *interpreted* as real numbers.

So we will try to show that all of mathematics can be done in the set–theoretic universe. Many of the set-theoretic ideas required are rather straightforward, and can be developed without recourse to the axiomatic method (though every mathematician must be aware of the potential pitfalls associated with a purely "naïve " view of sets).

But apart from its intrinsic importance as a fundamental vehicle for for doing *other* mathematics, set theory is also an independent branch of mathematics in its own right, regarded as one of the four main branches of mathematical logic. In other words, set theory — like group theory or number theory — is worth pursuing for its own sake: "Set theory is the mathematical science of the infinite", according to the set theorist Thomas Jech. You probably already know that the natural numbers $\mathbb{N}$ and the rational numbers $\mathbb{Q}$ both form *countable* sets — i.e. they have the same size (cardinality): $|\mathbb{N}| = |\mathbb{Q}|$. If you know that, you probably also know that the set of real numbers $\mathbb{R}$ is *uncountable*, and thus that infinity comes in different sizes, with $|\mathbb{N}| < |\mathbb{R}|$. Some historians date the birth of the autonomous subject *Set Theory* to that day (7 December 1873) that Georg Cantor — who is regarded as the creator of the subject — wrote to Richard Dedekind to inform him of this fact. A short while later, in 1878, Cantor framed the *Continuum Problem*: Is there a set A whose size is strictly between that of the natural numbers and the real numbers, i.e. such that $|\mathbb{N}| < |A| < |\mathbb{R}|$? Cantor was unable to settle this question, though he hypothesized that there is no such set A — the *Continuum Hypothesis*. The *über*–mathematician David Hilbert was one of the first to appreciate the power of set theory, and its importance to general mathematics. Accordingly the Continuum Problem was first on the famous and influential list of 23 problems that he presented at the Second International Congress of Mathematicians in Paris, 1900. This innocuous–appearing question initially resisted all efforts, and was finally settled in a rather surprising way: It was proved by Kurt Gödel in 1938 that the Continuum Hypothesis cannot be disproved, and then by Paul Cohen in 1963 that the Continuum Hypothesis cannot be proved! In order to obtain such results, it was necessary to write down a precise list of axioms for set theory — the Zermelo–Fraenkel Axioms — so that there would be agreement on what constitutes a valid set–theoretic proof. What Gödel and Cohen proved was that the Continuum Hypothesis is *independent* of these axioms. (There remains open the possibility that mathematicians of the future will discover a new plausible axiom for set theory, which will then settle the Continuum Problem in one way or the other[1].) Set theory is riddled with problems like these, where it

[1] Indeed, some mathematicians believe that the Continuum Hypothesis was recently settled by the set

can be proved that the problem is unsolvable[2]. Such independence proofs require methods from logic (model theory, proof theory, recursion theory), and thus an axiomatic approach is here indispensable.

This chapter is concerned with the foundations of set theory, and a little philosophy[3] will be unavoidable. In fact, the title of the first ever monograph dedicated to the subject — Cantor's *Grundlagen einer allgemeinen Mannigfaltigkeitslehre: Ein mathematisch–philosophischer Versuch in der Lehren des Unendlichen*, 1883 — makes clear that, in set theory, mathematics and philosophy were inextricably intertwined right from the start.

Exercise 1.1.1 All active branches of mathematics have problems that are unsolved. It is usually tacitly agreed that such an unproved statement is either true or false. But set theory has problems that are *provably unsolvable.* This leads to a philosophical question: Do you believe that the Continuum Hypothesis must be either true or false — we just do not know which? Or do you believe that it simply doesn't make sense to ask whether the Continuum Hypothesis is true or false? Can you back up your belief with an argument?

□

1.2 Naïve Set Theory

Classifying objects into sets is a basic operation of the human mind. It is more basic even than counting: Before you can count some objects, you first have to decide which objects to count, i.e. you first have to mentally group them together, and separate them from objects you do not want to count. Nothing seems more natural, therefore, than to collect objects which share a common property into a set. Naïve Set Theory (NST) is a formal theory which allows us to do just that. It has the following axioms:

NST 1. Extensionality: If two sets X, Y have the same elements, then $X = Y$.

NST 2.$_P$ **Full Separation:** If $P(x)$ is a property of sets, then there is a set

$$Y = \{x : P(x)\}$$

which consists precisely of those x which possess the property P.

Observe that the Full Separation Axiom is not a single axiom, but an *axiom schema*: We have an axiom for every possible property $P(\cdot)$ (though right now we won't be too precise about exactly what constitutes a "property"). Thus there are infinitely many axioms.

The first axiom states that a set is determined by its members, and by nothing else. In a sense, it tells us what it *means to be a set.* The second axiom allows us to show that many sets actually exist. For example:

theorist Hugh Woodin, and that it is false. Others dispute this.

[2] Assuming that the axioms of set theory are consistent, that is. But Gödel proved that we cannot prove that the axioms of a set theory are consistent without assuming the consistency of some more powerful theory. In other words, it is impossible to prove from the axioms of set theory that set theory is consistent.

[3] I highly recommend browsing the online *Stanford Encyclopedia of Philosophy*, `http://plato.stanford.edu` for good introductions to the philosphy of mathematics, and much more besides.

- If $P(x) \equiv (x \neq x)$, then $\{x : P(x)\}$ is a set with no elements (since $x = x$ for any x). By the Axiom of Extensionality any two empty sets are equal, because they have exactly the same elements. Hence a unique empty set exists, and we denote it by $\varnothing$.
- If $P(x) \equiv (x = x)$, then $\{x : P(x)\}$ is the set of all sets: Any set x is equal to itself.
- If $P(x) \equiv (x = a \vee x = b)$, then $\{x : P(x)\}$ is the set $\{a, b\}$.
- If $P(x) \equiv (\forall z \in x\ (z \in a))$, then $\{x : P(x)\}$ is the *power set* of a, i.e. the set of all subsets of a.
- If $P(x) \equiv (x \text{ is a real number} \geq 0)$, then $\{x : P(x)\}$ is the set of all real numbers greater or equal to zero. (At this point, we do not yet know that there are any real numbers, so it is at this stage possible that this set is empty.)

Towards the end of the 19th Century, mathematicians became increasingly more concerned with the foundations of mathematical knowledge, and the status of mathematical "reality". Some logicians — amongst whom were Gottlob Frege and Bertrand Russell — tried to reduce mathematics to logic. In June 1902, Russell wrote a letter to Frege, warmly expressing his appreciation for the latter's work, but adding that "There is just one point where I have encountered a difficulty". Russell asked Frege to consider the property

$$P(x) \equiv (x \notin x)$$

If $y = \{x : x \notin x\}$ exists (which, according to the Full Separation Axiom, it must), then there are two possibilities: Either $y \in y$ or $y \notin y$.

(i) If $y \in y$, then $P(y)$ must hold (since y consists of exactly those sets which have property P). In that case $y \notin y$ — contradiction.

(ii) If $y \notin y$, then certainly $P(y)$ holds. So y is one of the sets that has property P, and therefore $y \in \{x : P(x)\}$, i.e. $y \in y$ — contradiction.

Thus the theory NST is inconsistent! *Russell's paradox* had devastating consequences for Frege's programme to reduce arithmetic to logic.

Russell's paradox[4] (also called *Russell's antinomy*) is just the simplest example of the problems that may arise from a too liberal notion of set. Cantor himself was aware of such problems by 1899 at the latest, and dealt with this by denying the status of sethood to certain collections (multiplicities).

> *For a multiplicity can be such that the assumption that* all *of its elements 'are together' leads to a contradiction, so that it is impossible to conceive of the multiplicity as 'one finished thing'. Such multiplicities, I call* absolutely infinite *or* inconsistent multiplicities*. As we can readily see, the 'totality of everything thinkable', for example, is such a multiplicity; later still other examples will turn up.*
> *If on the other hand, the totality of the elements of a multiplicity can be though of without contradiction as 'being together', so that they can be gathered together into 'one thing', I call it a* consistent multiplicity *or* set*.*
>
> — Cantor, letter to Dedekind, 1899.

[4] Apparently, Ernst Zermelo was aware of Russell's paradox some years before. But it was Russell who understood its devastating consequences, and who brought it to the attention of the wider world.

Thus, according to Cantor, the 'totality of everything thinkable' is inconsistent: There is no *set of all sets*[5].

If we can't just lump anything thinkable together into a set, then what can we do? A number of different approaches have been suggested. Some mathematicians and philosophers thought that Russell's paradox indicated that there were problems not just with an intuitive notion of set, but with logic itself. (It is easy to rephrase Russell's paradox in purely logical terms[6]). The intuitionists, for example, thought that the basic laws of logic do not apply to infinite collections. Amongst others, they denied the validity of the *Law of the Excluded Middle*[7]. Other mathematicians thought that classical logic was inherently sound, and that one need merely analyse the notion of set with greater care and precision, if one wants to avoid contradictions and inconsistencies. This is the approach we will take. But avoiding an inconsistent set theory is harder than one might think. Firstly, a number of set theories have been proposed with the aim of avoiding Russell's paradox, but which subsequently turned out to be inconsistent after all — notably the system ML (Mathematical Logic) proposed in 1940 by the great logician Willard van Orman Quine. And many of these set theories are artificial: They are, as Russell commented, not "such as even the cleverest logician would have thought of if he had not known of the contradictions."

In the next section, we will look at an intuitively compelling view of the set–theoretic universe, in which sets are built up in stages, with each set made up from simpler sets. This picture of the set–theoretic universe then suggests certain axioms for set formation, the *Zermelo–Fraenkel* Axioms, or ZF. If ZF set theory is consistent, then we may never know it: Gödel's 2^{nd} Incompleteness Theorem states (roughly) that a theory that is strong enough to contain basic arithmetic cannot prove its own consistency. If ZF is inconsistent, then one day we may find a contradiction. However, ZF has been used for nearly a century already, and none has been found as yet[8].

[5] How did Cantor come to this conclusion? Define the *power set* $\mathcal{P}(X)$ to be the set of all subsets of X. Cantor proved that $|X| < |\mathcal{P}(X)|$. Now if X is the set of *all* sets, then $\mathcal{P}(X) \subseteq X$, since every member of $\mathcal{P}(X)$ is a set. But then $|\mathcal{P}(X)| \leq |X|$, and hence the assumption that all sets can be collected into a single set–of–all–sets leads to contradiction. Cantor knew about this by 1899 at the latest, though it was only published in 1932.

[6] Here's how: "It seems to make perfect sense to inquire of a property whether it applies to itself or not. The property of being red, for instance, does not apply to itself, since red is surely not red, whereas (the property of being) abstract, being itself abstract, applies to itself. Calling the property of not applying to itself 'impredicable', we arrive at the paradoxical consequence that impredicable is impredicable if and only if impredicable is not impredicable. The property–theoretical (logical) variant is as paradoxical as the set–theoretical (mathematical) one." — *Foundations of Set Theory*, by Fraenkel, Bar–Hillel and Levy, 2nd revised edition, Springer 1973.

[7] This is the law which asserts that for any statement A, either A is true or A is false. The intuitionists therefore denied that the tautology $A \vee \neg A$ is true.

[8] In September 2011, the mathematician Edward Nelson announced that arithmetic (to be precise, *Peano Arithmetic* (PA)) is inconsistent. His proof turned out to have an error. Nelson is hardly a "lightweight" mathematician, and this episode reveals that serious mathematicians may have legitimate qualms about the consistency of even the most basic and longstanding mathematical theories.

By Gödel's 2^{nd} Incompleteness Theorem, PA is not strong enough to prove its own consistency. ZF is strong enough to prove that PA is consistent ($ZF \vdash \text{Con}(PA)$), but since we don't know if ZF is consistent, that proof is also suspect. (In 1936 the logician Gerhard Gentzen proved that PA is consistent. His proof goes beyond ordinary arithmetic by requiring a *transfinite* induction, but only up to a countable ordinal.)

1.3 The Iterative Conception of Set

> **Definition:** *By a "set" we mean any collection M into a whole of definite distinct objects m (which are called the "elements" of M) of our perception or of our thought.*
>
> — Georg Cantor, 1895
>
> (First sentence of *Contributions to the Founding of the Theory of Transfinite Numbers.*)

> *A set, according to Cantor, is 'any collection ..., into a whole of definite, well-distinguished objects... of our intuition or thought.' Cantor also defined a set as a 'many, which can be thought of as one, i.e., a totality of definite elements that can be combined into a whole by a law.' One might object to the first definition on the grounds that it uses the concepts of collection and whole, which are notions no better understood than that of set, that there ought to be sets of objects that are not objects of our thought, that 'intuition' is a term laden with a theory of knowledge that no one should believe, that any object is 'definite,' that there should be sets of ill-distinguished objects, such as waves and trains, etc., etc. And one might object to the second on the grounds that a many' is ungrammatical, that if something is a many'' it should hardly be thought of as one, that totality is as obscure as set, that it is far from clear how laws can combine anything into a whole, that there ought to be other combinations into a whole than those effected by laws,' etc., etc.*
>
> — George Boolos, 1971
>
> (Opening lines *The Iterative Conception of Set*,
> J. Philosophy 68 no.8, 1971, pp. 215-231.)

Russell's paradox involved the multiplicity of all sets which are not elements of themselves. Can a set belong to itself?

Looked at one way, the answer appears to be "Yes!": Let A be the set of all sets definable, in English, using fewer than 100 words. We have just defined A in English using fewer than 100 words. Hence $A \in A$. However, "definable in English" is a rather obscure term. English has just finitely many words, so there are just finitely many sentences of length < 100 words. Thus there are at most finitely many natural numbers definable in < 100 words. Now let n be the smallest natural number *not* definable in fewer than 100 words.... Contradiction...

Another example is the set

$$X := \{\{\{\ldots \varnothing \ldots\}\}\}$$

where there are infinitely many brackets arount the empty set. Then clearly $\{X\} = X$, so $X \in X$.

But looked at from another angle, the answer to our question appears to be "No!", however: In order to collect some objects into a set — i.e in order to *make* a set — one has to *have* those objects first. Thus the elements which make up a set are, in some sense, more primitive than the set itself. As an analogy, one can think of the natural numbers as being "created" from the number 0 by the operation of repeatedly adding 1. We start with zero, then we make

1, then we make 2, then we make 3, etc. Now it must not be thought that natural numbers are continuously being created — "make" is just a manner of speaking. In the same way, one can think of sets of being made up as follows: We start off with some primitive objects called *atoms* (also called *individuals*, or *urelemente*), which are *not* sets, and have no elements themselves. At stage 0, we construct all sets of atoms. Then at stage 1, we construct all sets made up of atoms and sets created at stage 0. At stage 2, we construct all sets consisting of atoms and sets created at stages 0 and 1. Thus if we start with two atoms a, b, then we get the following:

Stage 0: New sets $\varnothing, \{a\}, \{b\}, \{a,b\}$.

Stage 1: New sets $\{\varnothing\}, \{\{a\}\}, \{\{b\}\}, \{\{a,b\}\}, \{\varnothing, \{a\}\}, \{\varnothing, \{b\}\}, \{\varnothing, \{a,b\}\}, \{\varnothing, \{a\}, \{b\}\}, \dots$

Stage 2: New sets $\{\{\varnothing\}\}, \{\varnothing, \{\varnothing\}\}, \{\{\{a\}\}\}, \dots$

Stage 3: $\dots$

We do this for all the stages 0,1,2,.... But it doesn't stop there! Once we have all these sets, we can construct sets made up of atoms and sets constructed at some finite stage. Thereafter we can construct sets made up of atoms and sets just constructed, at the first infinite stage, for example the set

$$\Big\{\varnothing,\ \underbrace{\{\varnothing\}}_{\text{1 bracket}},\ \underbrace{\{\{\varnothing\}\}}_{\text{2 brackets}},\underbrace{\{\{\{\varnothing\}\}\}}_{\text{3 brackets}},\dots\Big\} \qquad (\star)$$

You may have noticed that we have not said anything about the nature of the atoms. What should or could atoms be? We might take them to be basic objects that we need in mathematics, e.g. natural – or real numbers, or points and lines. However, it turns out that we do not need atoms at all! Instead, we can build up all of mathematics from *nothing*, i.e. from the empty set. To do this, define a *pure set* as follows:

- Atoms are not pure.
- A set is pure if and only if all its elements are pure.

Thus at stage 0, we form the set $\varnothing$, which is pure — vacuously, because none of its members are impure. At stage 1, the set $\{\varnothing\}$ is pure. At stage 2, we have the pure sets $\{\{\varnothing\}\}$ and $\{\varnothing, \{\varnothing\}\}$, etc. The infinite set $(\star)$ constructed at the first infinite stage is clearly pure as well.

We will *only* be concerned with *pure sets*. Thus when we say "set", we mean "pure set". This means that in our mathematical universe, *everything is a set*.

Now observe that sets are made again and again. For example, $\varnothing$ is a subset of all sets, and hence is created anew at each stage. But for each set there is a least stage, i.e. a stage at which the set appears for the first time.

We thus have the following intuitions about set formation via stages:

(i) There is a least stage 0 at which we have the set $\varnothing$.

(ii) Every stage α has a successor $\alpha+1$. At stage $\alpha+1$ we can construct sets whose elements are sets constructed at stage α and earlier. This means that if a collection (multiplicity) has elements all of which are created at stage α or earlier, then that collection is a set, and is "created" by stage $\alpha+1$.

(iii) Having made sets at a whole lot of stages, there is least a new stage, at which we can construct sets whose elements are sets constructed at an earlier stage.

(iv) A collection (multiplicity) is a set if and only if it is formed at some stage.

(v) For every set, there is a least stage at which that set is formed.

Thus we have stages $0, 1, 2, 3, \ldots, n \ldots$. After that comes a stage — call it stage ω. After stage ω come stages

$$\omega + 1, \omega + 2, \ldots, 2\omega, 2\omega + 1, \ldots, 3\omega, \ldots, \omega \cdot \omega, \ldots, \omega^3, \ldots, \omega^\omega, \ldots, \omega^{\omega^\omega}, \ldots, \omega^{\omega^{\omega^{\omega^{\vdots}}}}, \ldots$$

Thereafter comes another stage — call it ω_1, etc. These stages conform to Cantor's two "generating principles[9]": (1) Every stage α has an immediate successor stage $\alpha + 1$, and (2) given any sequence of stages without a last element, there is a stage which comes immediately afterwards.

If α, β are stages, we write $\alpha < \beta$ to mean that stage α comes before stage β. When we form a set at some stage β, its elements must be sets that have been constructed at some stage $\alpha < \beta$. It is therefore impossible for a set to belong to itself. Furthermore, it is impossible to get a loop such as

$$x_1 \in x_2 \in x_3 \cdots \in x_n \in x_1$$

Now we consider a property that will turn out to play a very important role in the future development of set theory: Let Δ be any non–empty collection of stages, and let Γ be the collection of all stages that come strictly before any of the stages in Δ, i.e.
$\Gamma := \{\gamma : \forall \delta \in \Delta \ (\gamma < \delta)\}$. We can then consider two cases:

- Either Γ is empty. In that case if the zeroth stage 0 belongs to Δ, so 0 is the least element of Δ.

- Else Γ is non–empty. In that case intuition (iii) above says that there is a least stage α which lies above all $\gamma \in \Gamma$. We now claim that α is the least element of Δ. To see this, note first that each $\delta \in \Delta$ lies strictly above all $\gamma \in \Gamma$. Since α is the *least* stage which lies above all $\gamma \in \Gamma$, it follows that $\forall \delta \in \Delta \ (\alpha \leq \delta)$. Next observe that it cannot be the case that $\forall \delta \in \Delta \ (\alpha < \delta)$, for then $\alpha \in \Gamma$. Hence there is some $\beta \in \Delta$ such that $\beta \leq \alpha$. But since $\beta \in \Delta$ we also have $\alpha \leq \beta$, and thus $\alpha = \beta$. In particular, $\alpha \in \Delta$. Since $\forall \delta \in \Delta \ (\alpha \leq \delta)$, we see that α is the least element of Δ.

In both cases, therefore, we have shown that Δ has a least element:

Any non–empty collection of stages has a least element.

All of the above is at the *intuitive level*, and must be made formal. And we will accomplish this: For example, the intuitive idea of a *stage* will be captured by the formal notion of an *ordinal*. The fact that each set is "created" at some stage will be be evident when we consider the *cumulative hierarchy of sets* and the *Axiom of Foundation*.

[9] Cantor, *Grundlagen einer allgemeinen Mannichfaltgkeitslehre*, 1883.

1.4 The Axioms of ZFC

> *Set theory is that branch of mathematics whose task it is to investigate mathematically the fundamental notions "number", "order", and "function", taking them in their pristine, simple form, and to develop thereby the logical foundations of all arithmetic and analysis; thus it constitutes an indispensable component of the science of mathematics. At present, however, the very existence of the discipline seems to be threatened by certain contradictions, or "antinomies", that can be derived from its principles — principles necessarily governing our thinking, it seems – and to which no entirely satisfactory solution has yet been found.[...]Under these circumstances there is at this point nothing left for us to do but to proceed in the opposite direction, and starting from set theory as it is historically given, to seek out the principles required for establishing the foundations of this mathematical discipline. In solving the problem we must, on the one hand, restrict these principles sufficiently to exclude all contradictions and, on the other, take them sufficiently wide to retain all that is valuable in this theory.*
> *Now in the present paper I intend to show how the entire discipline created by Cantor and Dedekind can be reduced to a few definitions and seven principles, or axioms, which appear to be mutually independent.*
>
> — Ernst Zermelo, 1908
>
> (Opening lines of *Untersuchungen über die Grundlagen der Mengenlehren I*, Mathematischen Annalen 65, pp. 261-281, 1908. Translation in van Heijenoort, *From Frege to Gödel*)

You may already have encountered various systems of axioms, such as the axioms for a vector space, or of a field. A *model* of a system of axioms is a mathematical object that satisfies these axioms. Thus a field is a model of the field axioms.

Systems of axioms can be separated into two classes:

I. Some systems of axioms are obtained by studing a number of mathematical objects/theories and *abstracting* certain common elements. Examples of such are the axioms of group theory, of field theory, of vector spaces, of topology, of lattice theory, of probability theory.... These systems of axioms have many models, i.e. there are many different kinds of group, many different kinds of field, many different kinds of vector space, ... Moreover, these axioms are intended to have many models.

II. Other systems of axioms are meant to describe and characterize one kind of mathematical objects, i.e. these axioms have an intended model. The axioms of arithmetic are meant to capture just the set $\mathbb{N}$ of natural numbers with the operations $+, \times$ and relation $\leq$ on $\mathbb{N}$. The axioms of a complete ordered field are intended to capture the set of real numbers. Euclid's axioms of geometry are intended to characterize space.

    In some cases, systems of axioms succeed in characterizing exactly the intended model. Notably the axioms for a complete ordered field exactly capture the reals, in that any two models of these axioms are isomorphic (i.e. essentially "the same"). In many cases, it is impossible to completely capture an object with a set of axioms: If a set of first–order axioms has just one an infinite model, it will have many non–isomorphic models[10].

The axioms of set theory are of the second kind: They are intended to describe the *universe of all sets*. Below, we will write down these axioms, and briefly justify them in terms of the iterative conception of sets. In the chapters that follow, we will develop the consequences of these axioms in some detail, and (hopefully) convince you that all of mathematics can be interpreted in the universe of sets.

The first set of axioms for set theory were presented by Ernst Zermelo in 1908, in a paper entitled *Investigations in the Foundations of Set Theory I.* An additional schema of axioms was suggested by Abraham Fraenkel in 1922 (and independently, by Thoralf Skolem in 1923). We shall phrase these axioms in the language of first order predicate logic. The language of set theory has at its disposal the following symbols, whose meaning is assumed to be already known to you. We have at our disposal a countable collection of

- Variables: $v_0, v_1, v_2, v_3, \ldots$

As non–logical symbols, we have the two binary predicate symbols

- Equality: $=$
- Membership: $\in$

The atomic formulas of our theory are expressions

$$x = y \qquad\qquad x \in y$$

and these can be combined into formulas by means of

- Logical connectives: $\wedge$ (conjunction, and),$\vee$ (disjunction, or),$\neg$ (negation, not), $\rightarrow$ (implication, implies), $\leftrightarrow$ (equivalence, if and only if).
- Quantifiers: $\forall$ (for all), $\exists$ (there exists).
- Brackets: (,)

It is assumed that you already understand what a (well–formed) formula is, when an instance of a variable in a formula is *free* or *bound*, etc. All of this is part of our *metatheory*, the basic background theory in which we discuss our *formal* set theory. It's obvious — if you stop to think for a minute — that one cannot develop any theory (mathematical, or otherwise) from absolutely nothing. We have to assume that we have a common background language in which to develop our formal theory, and a common basis of knowledge that we can use as springboard from which to start. For us, this *metalanguage* and *metatheory* includes a natural language (English) augmented by a few mathematical notions, such as the intuitive notion of natural number, the notion of a finite set, and very elementary first–order logic. This foundation, involving only intuitively obvious *finitary* notions and – reasoning, is (hopefully) obvious *a priori*, and its validity not in doubt[11].

[10] This is the *Löwenheim–Skolem Theorem*, a basic result in the branch of mathematical logic called *model theory.* It follows from this theorem that if the axioms of set theory are consistent, then there is a *countable* model of set theory, i.e. a model in which there re just countably many sets — a result that Skolem found highly paradoxical.

[11] If you *do* doubt it, your mathematics will look very different. For example, for the intuitionists *every* function $f : \mathbb{R} \rightarrow \mathbb{R}$ is continuous. This follows logically from certain assumptions that are quite natural.

Apart from the above symbols, we will introduce symbols for defined notions as we proceed. Many of these will undoubtedly be familiar to you, such as $\cup, \cap, \varnothing, -$ for union, intersection, empty set and set difference. We will use $x, y, z, X, Y, Z, \ldots$ as symbols for variables, and also use brackets $[,], \Big(,\Big), \Big[,\Big]$ to make expressions easier to understand. We will also employ expressions such as $x \neq y$ rather than the formally correct $\neg(x = y)$, and we will omit brackets wherever it makes an expression easier to read without risk of ambiguity. Thus, for example, we will write

$$\forall x\ \exists y\ (x \notin y) \quad \text{instead of} \qquad \forall v_0\ (\exists v_1\ (\neg(v_0 \in v_1)))$$

The latter is a well–formed sentence of our formal language, but the former is much easier to read.

We also assume that you are aware of various logical equivalences, such as

$$\varphi \vee \psi\ \leftrightarrow\ \neg(\neg\varphi \wedge \neg\psi) \qquad \text{and} \qquad \forall x\ \varphi\ \leftrightarrow\ \neg\exists x\ (\neg\varphi)$$

etc.

Furhermore, we introduce the following notation:

$$(\forall x \in y)\ \varphi \qquad \text{means} \qquad \forall x(x \in y\ \rightarrow\ \varphi)$$

and

$$(\exists x \in y)\ \varphi \qquad \text{means} \quad \exists x\ (x \in y\ \wedge\ \varphi)$$

ZFC 0: The Axiom of Set Existence

ZFC 0:
There is a set.

$$\exists x\ (x = x)$$

Without this axiom, we have nothing to talk about. (For some systems of axioms, models with no elements are acceptable, and even desirable. Category theorists, for example, like to work with categories that have *initial objects*, and these can be empty. But our axioms are meant to characterize a single "object": The universe of all sets — and we believe that that universe is non–empty.) This axiom is not strictly necessary, as it is implied by the Axiom Infinity.

ZFC 1: The Axiom of Extensionality

ZFC 1:
If two sets have the same elements, then they are equal.

$$\forall x\ \forall y\ [\forall z\ (z \in x \leftrightarrow z \in y) \rightarrow x = y]$$

The *extension* of a property is the collection of all objects which have that property, i.e. to which that property extends.

It is not *a priori* clear that

$$\{\text{Moons of Venus}\} = \{\text{Moons of Mercury}\}$$

or that

$$\{x^2 : x \in \mathbb{R}\} = \{y^{10} : y \in \mathbb{R}\}$$

The *intension* of those definitions of sets — i.e. the *way* in which they are defined — is different. However their *extension* is the same. Both the sets {Moons of Venus} and {Moons of Mercury} are empty, and are thus equal to $\varnothing$. Both the sets $\{x^2 : x \in \mathbb{R}\}$ and $\{y^{10} : y \in \mathbb{R}\}$ consist exactly of all non–negative real numbers, and are thus equal to $[0, \infty)$. The Axiom of Extensionality states the *intension* is not important when it comes to set equality, but only the *extension.*

We may now introduce a new symbol $\subseteq$ for the relation of *inclusion*, a binary relation on sets which is defined as follows.

Definition 1.4.1 We say that a set x is a *subset* of y and write $x \subseteq y$ if and only if every element of x is an element of y:

$$x \subseteq y \qquad \Longleftrightarrow \qquad \forall z\ (z \in x \to z \in y)$$

We also write $y \supseteq x$ instead of $x \subseteq y$.

We say that x is a *proper subset* of y, and write $x \subsetneq y$ (or $y \supsetneq x$), if $x \subseteq y$ and $x \neq y$.

□

The Axiom of Extensionality can now be written thus:

$$x \subseteq y \wedge y \subseteq x \to x = y$$

Remarks 1.4.2

- The binary relations $\in$ and $\subseteq$ are sometimes confused, especially by beginners. They mean quite different things, however. Indeed, observe that it is always the case that $x \subseteq x$, but that it is seldom the case that $x \in x$. (Indeed, never: That $x \in x$ contradicts the iterative conception of sets was discussed in §1.3, and will follow formally from the Axiom of Foundation.)
- Another role of the Axiom of Extensionality is to limit the universe of discourse to sets. In particular, it says that if two objects x, y of our universe have the same elements, then $x = y$. It therefore follows, for example, that there are no atoms. (For if a is an atom, then a and $\varnothing$ have the same elements, so $a = \varnothing$ — contradiction.)

□

ZFC 2: The Axiom Schema of Separation

The remaining axioms, apart from the Axiom of Foundation, are axioms of *set formation*, i.e. they assert that certain sets exist, or that certain collections (multiplicities) are sets. The guiding principle, expounded by Cantor, Russell and Zermelo, is that of *limitation of size.* Sets are collections that are "not too big." This prevents one forming a "set of all sets", for example.

ZFC 2_φ:
Let $\varphi(x, p_1, \ldots, p_n)$ be first–order formula with free variables amongst $x, p_1, \ldots, p_n$. For any set z, there is a set y consisting of all those $x \in z$ which have the property φ.

$$\forall z\ \forall p_1\ \ldots\ \forall p_n\ \exists y\ \forall x\ (x \in y \leftrightarrow x \in z \wedge \varphi(x, p_1, \ldots, p_n))$$

Given a set z and a property φ, we may *separate* all those elements of z which have the property φ, and put them together to form a new set y. By the Axiom of Extensionality, there is just one such set y. We therefore may introduce some new notation: Let

$$\{x \in z : \varphi(x,p)\}$$

denote the unique set y whose existence is asserted by the Axiom Schema of Separation. Observe that y is a sub–collection of z. The Separation Axiom asserts that the collection (multiplicity) y is a set.

It is easy to justify the existence of such a set y from the iterative conception of sets: Given the set z, we know that it is "created" at some stage α. All the elements x of z are therefore created at some stage strictly earlier than α. In particular, the elements of y are created at some stage strictly earlier than α, as $y \subseteq z$. But at stage α we make sets out of *all* those collections whose elements are sets created at an earlier stage. In particular, we make a set out of the collection y at (or possibly even before) stage α.

Observe that Separation is not a single axiom, but an *axiom schema*: We have one axiom for every first–order $\varphi(x,p)$. However, given any formula (or string of symbols) Φ, it is a mechanical procedure to check whether or not Φ is an instance of the separation axiom.

The Axiom Schema of Separation is also called the *Axiom Schema of Comprehension* or the *Aussonderung–axiom* in the literature. It will, in fact, follow from the Axiom Schema of Replacement (ZFC 7), so it isn't strictly necessary. But it is easier to use than Replacement, and is on the original list of axioms that Zermelo published in 1908.

Exercise 1.4.3 (a) Prove that the existence of a unique empty set follows from the axioms ZFC 0–2. We may therefore define $\varnothing$ to be *the unique* set with no elements.

(b) Prove that ZFC 0–2 imply that every set x has at least one subset y which is not an element of x, i.e. prove that $\forall x\ \exists y\ (y \subseteq x \wedge y \notin x)$. Deduce that the collection of all sets is not a set.

(c) Recall that if a, b are sets, then their *intersection* $a \cap b$ is the collection of all elements which belong to both a and b. Further recall that $a - b$ is the collection of all elements which belong to a but not to b. Show that ZFC 0–2 imply that if a, b are sets, then so are $a \cap b$ and $a - b$.

□

Remarks 1.4.4 In the exercise above, we formally introduced the symbols $\varnothing, \cap, -$. Previously, we introduced the symbol $\subseteq$ and the notation $\{x \in a : \varphi\}$. We will use these *introduced notions* as though they are part of our formal language. They can always be eliminated, however. For example,

$$\forall x\ \exists y\ (y \subseteq x) \qquad \text{is shorthand for} \qquad \forall x\ \exists y\ \forall z\ (z \in y \to z \in x)$$

and

$$\{y \in a : \varphi(y,p)\} \cap b = \varnothing \qquad \text{is shorthand for} \qquad \exists z \left[\forall y\ (y \in z\ \leftrightarrow\ y \in a \wedge \varphi(y,p)) \wedge \neg \exists x\ (x \in z \wedge x \in b)\right]$$

□

ZFC 3: The Axiom of Pairing

ZFC 3:
Given two sets x, y there is a set z whose members are precisely x, y.

$$\forall x\ \forall y\ \exists z\ \forall w\ (w \in z \leftrightarrow w = x \vee w = y)$$

By the Axiom of Extensionality, any two sets whose members are precisely x, y are equal, and thus we may define the *unordered pair* $\{x, y\}$ to be the unique set whose elements are precisely x and y, i.e.

$$z = \{x, y\} \quad \text{is shorthand for} \quad x \in z\ \wedge\ y \in z\ \wedge\ \forall w\ (w \in z\ \rightarrow\ w = x\ \vee\ w = y)$$

We then define the singleton $\{x\}$ to be the set $\{x, x\}$.

The Axiom of Pairing is easy to justify via the iterative conception of sets: If a and b are sets, then each is created at some stage, say stages α and β. Then $\{a, b\}$ will exist at any stage γ which is strictly greater than both α and β.

Remarks 1.4.5 The Axiom of Pairing is often stated in the following weaker form: Given any two sets x, y, there exists a set z that they both belong to. However, together with the Separation axiom, this weaker form implies our form. For if z is a set which has x, y amongst its elements, then

$$\{x, y\} := \{w \in z : w = x \vee w = y\}$$

□

ZFC 4: The Axiom of Union

ZFC 4:
For any set x, there is a set y whose elements are precisely the elements of elements of x.

$$\forall x\ \exists y\ \forall z\ [z \in y\ \leftrightarrow\ \exists w \in x\ (z \in w)]$$

By the Axiom of Extensionality, there is a unique such set y, which is called the *union* of the set x, and denoted by

$$y = \bigcup x$$

or (more commonly) by

$$y = \bigcup \{w : w \in x\} \qquad \text{or} \qquad y = \bigcup_{w \in x} w$$

We can justify this axiom as follows: If x is a set, then it is formed at some stage α. Each element of x is therefore created at some stage $< \alpha$, and thus the elements of elements of x are created before stage α as well. Now the union of x — i.e. the collection of all elements of elements of x — is a collection of elements that are all created strictly before stage α, and hence $\bigcup x$ is formed as a set by stage α at the latest.

Remarks 1.4.6 The Axiom of Union may be stated in the following weaker form: Given any set x, there exists a set z that has amongst its elements all the elements of elements of x. Together with the Separation axiom this weaker form implies our form. For if z such a set, then

$$\bigcup x := \{y \in z : \exists w \in x\ (y \in w)\}$$

□

We may now define (with the aid of the Axiom of Pairing) the sets

$$x \cup y := \bigcup\{x, y\} \qquad x \cup y \cup z := \bigcup\{x, y, z\} \qquad x_1 \cup \cdots \cup x_n = \bigcup\{x_1, \ldots, x_n\}$$

etc.

Exercise 1.4.7 (a) Suppose that $\mathbf{C} := \{y : \varphi(y, p)\}$ is the collection of all sets that have some property[12] given by the first–order formula φ. $\mathbf{C}$ need not be a set itself — it might be "too big" — but each of its elements is a set. Define the intersection $\bigcap \mathbf{C}$ of $\mathbf{C}$ to be the collection of all sets z with the property that z belongs to each element of $\mathbf{C}$, i.e.

$$\bigcap \mathbf{C} = \{z : \forall y \in \mathbf{C}\ (z \in y)\}$$

Show that ZFC 0–2 imply that if the collection $\mathbf{C}$ is non–empty, then $\bigcap \mathbf{C}$ is a set. Further, use ZFC 0–3 to define the intersections $x \cap y$, $x \cap y \cap z$ etc. Observe that the Axiom of Union is not required for the existence of intersections.

(b) Show that $\bigcap \varnothing$ is not a set.

□

We may also write $y = \bigcap C$ as $\bigcap\{y : y \in C\}$ or $y = \bigcap_{y \in C} y$. Further, $x \cap y := \bigcap\{x, y\}$, etc.

Exercise 1.4.8 If the following identities and inclusions are unfamiliar, *now*[13] is the time to prove them.

(a) $x \subseteq y \cap z$ iff $x \subseteq y$ *and* $x \subseteq z$; $x \supseteq y \cup z$ iff $x \supseteq y$ *and* $x \supseteq z$.

(b) $x \subseteq y$ iff $x \cap y = x$ iff $x \cup y = y$ iff $x - y = \varnothing$.

(c) The *distributive laws*: $x \cap (y \cup z) = (x \cap y) \cup (x \cap z)$ and $x \cup (y \cap z) = (x \cup y) \cap (x \cup z)$.

(d) *De Morgan's Laws*: $x - (y \cup z) = (x - y) \cap (x - z)$ and $x - (y \cap z) = (x - y) \cup (x - z)$.

ZFC 5: The Axiom of Power Set

ZFC 5:
For any set x, there is a set y whose elements are precisely the subsets x.

$$\forall x\, \exists y\, \forall z\, (z \in y \ \leftrightarrow\ z \subseteq x)$$

By the Axiom of Extensionality, there is a unique such set y, which is called the *power set* of the set x, and denoted by

$$y = \mathcal{P}(x)$$

The Axiom of Power Set may be justified as follows: If x is a set formed at stage α, then all the elements of x are formed before stage α. But then each subset of x is a collection of elements that are created before stage α, so each subset of x is made into a set by stage α at the latest. So the collection of all subsets of x is a collection of sets that all exist by stage α, and thus $\mathcal{P}(x)$ is made into a set by stage $\alpha + 1$ at the latest.

[12] The technical term for such a collection is *class* — cf. §1.5.

[13] Yes, NOW!

Exercise 1.4.9 Show that if x, y are sets, then $\{x, y\} \in \mathcal{P}(\mathcal{P}(x) \cup \mathcal{P}(y))$.

Given the above, do we need the Axiom of Pairing? Or does Pairing follow from Union and Power Set?

□

ZFC 6: The Axiom of Infinity

> *"But concerning your proof, I protest above all against the use of an infinite quantity as something completed, which is never permissible in mathematics. The infinite is merely a way of speaking, the true meaning being a limit which certain ratios approach indefinitely close, while others are permitted to increase without restriction."*
>
> — Gauss, letter to H. Schumacher, 1831

> *" The existence of the infinite will never again be deniable while that of the finites is nevertheless upheld. If one permits either to fall, one must do away as well with the other."*
>
> — Cantor, *Grundlagen*, 1883.

The axioms ZFC 0 – 5 allow us to prove that there are infinitely many sets, but they do not allow us to prove that there is an infinite set. This is not hard to see: The axioms ZFC 0 – 2 allows us to deduce the existence of possibly just one set, namely $\varnothing$. Further, axioms ZFC 2 – 5 allow us to create new sets from old, by taking subsets, pairs, unions, and power set. But if we have at our disposal only finite sets (i.e. if we are at a stage where only finite sets have been created), then the operations of taking subsets, pairing, union and power set are not capable of forming infinite sets. Indeed, even ZFC 0–5 together with ZFC 7–9 are not sufficient to prove the existence of an infinite set, and the existence of such a set is, in fact, *denied* by the *intuitionists* and *finitists*. Following Aristotle, the concept of infinity was separated into two distinct versions. On the one hand, there is the idea of an *actual* infinity, i.e. a completed, finished, definite thing consisting of infinitely many elements or parts. On the other, there is *potential* infinity, where one has a sequence of things which is never completed or finished. Aristotle allowed only *potential* infinities as non–paradoxical, and many mathematicians and philosophers agreed. Certainly, this point of view, eloquently expressed in the quote of Gauss above, had much to recommend it, and under its guidance mathematicians were able to develop completely rigorous definitions of limits without the use of infinitesimals[14].

It was Cantor who made the use of the *actual* infinity respectable in mathematics. It was Zermelo who first understood that its existence had to be *asserted*:

ZFC 6:
There is an infinite set.

$$\exists x \, [\varnothing \in x \ \wedge \ \forall y \, (y \in x \ \rightarrow \ y \cup \{y\} \in x)]$$

[14] The philosopher George Berkeley ridiculed the use of infinitesimals: "May we not call them the ghosts of departed quantities?" Strangely enough, Cantor too was resolutely opposed to the use of infinitely small numbers, and "refuted" them with arguments that could be applied equally to his infinitely large transfinite numbers. But the logician Abraham Robinson resurrected them and legitimised their use with his invention of the hyperreals and non–standard analysis.

Let us analyze this axiom: Define an operation $S(\cdot)$ on sets by

$$S(x) := x \cup \{x\}$$

This means that

$$y = S(x) \iff \forall z\ (z \in y \leftrightarrow z \in x \vee z = x)$$

By the Axioms of Pairing and Union, $S(x)$ is a set whenever x is. For reasons that will become clear later, the function $S(\cdot)$ is called the *successor function.* The Axiom of Infinity states that there is a set which has the following amongst its elements:

$$\varnothing,\ S(\varnothing),\ S(S(\varnothing)),\ S(S(S(\varnothing))), \dots$$

"Clearly", such a set must be infinite. However, we have not yet defined what we mean by an infinite set — this requires the formal notion of a natural number. Let us say that a set x is *inductive* if it satisfies the condition in the Axiom of Infinity, i.e. if

$$\varnothing \in x\ \wedge\ \forall y\ (y \in x\ \rightarrow\ S(y) \in x)$$

The Axiom of Infinity therefore asserts that there is an inductive set. (The set of natural numbers will then be defined to be the smallest inductive set.) Once we have defined what we mean by an infinite set, we shall be able to prove that an inductive set is indeed infinite. Moreover, we shall show that if an infinite set exists, then so does an inductive set. Thus the existence of an infinite set will turn out to be equivalent to the existence of an inductive set.

Let us quickly see how to justify this axiom via the iterative conception of sets: Observe that $\varnothing$ is formed at stage 0. Then $S(\varnothing) = \{\varnothing\}$ is a set whose elements are formed at stage 0, so $S(\varnothing)$ is formed at stage 1. Then $S(S(\varnothing)) = \{\varnothing, \{\varnothing\}\}$ is a set whose elements are formed at stage 1 and earlier so $SS(\varnothing))$ is made into a set at stage 2. Thus:

$$\begin{array}{rl}
\varnothing & \text{is formed at stage } 0 \\
S(\varnothing) & \text{is formed at stage } 1 \\
SS(\varnothing) & \text{is formed at stage } 2 \\
SSS(\varnothing) & \text{is formed at stage } 3 \\
 & \vdots
\end{array}$$

These sets therefore are all formed *before* the first infinite stage ω, and hence the inductive set

$$\{\varnothing, S(\varnothing), SS(\varnothing), SSS(\varnothing), \dots\}$$

is formed at stage ω.

ZFC 7: The Axiom of Choice

ZFC 7:
Every family of non–empty sets has a *choice function*: If x is a family of non–empty sets, then there is a function f on x which *chooses* from each $w \in x$ an element $f(w) \in w$.

$$\forall x\ [\forall w \in x\ (w \neq \varnothing) \rightarrow \exists f\ (\ f \text{ is a function} \wedge \mathrm{dom}(f) = x \wedge \forall w \in x\ (f(w) \in w))]$$

In the next chapter, we will see that functions can be interpreted in the universe of sets, so that to the statement "f is a function" can be written in the language of set theory. Observe that if f is a choice function on x, then $f : x \to \bigcup x$. (Of course, not every $f : x \to \bigcup x$ is a choice function.)

The Axiom of Choice — formulated explicitly by Zermelo in 1904 — is an immensely powerful tool. It seems intuitively obvious, and yet has generated so much controversy that it deserves (and will get) a chapter on its own. The reason that this axiom was initially eyed with such suspicion by a large number of mathematicians is that the axiom is *non–constructive.* It simply *asserts* the existence of an object — a choice function — without saying *how* such an object might be constructed.The axioms ZFC 2–6 have a slightly different character: They assert that certain *well–defined* collections are *sets*. The Axiom of Choice (abbreviated AC) , on the other hand, asserts the existence of an *undefined* set[15]. And this is precisely the source of its power: If one could *define* a choice function for a particular family x, then one would not need the AC to assert its existence.

Remarks 1.4.10 It is important to note that one does *not* need AC to "choose" an element from a non–empty set. This fact has often been misunderstood. The fact is that AC doesn't really allow you to "choose"; that's just a manner of speaking. What AC *does* do is assert that a certain set exists. Thus if w is non–empty set, then $x := \{w\}$ is family of non–empty sets, and AC says that there is a function $f : x \to \bigcup x$ such that $f(w) \in w$. But we don't need AC for that ! If w is non–empty, then (by definition of "non–empty" and first–order logic) there is an element $a \in w$. We may not know what a looks like — i.e. we may not explicitly be able to choose a particular a — but there *is* such an a. It will be clear after you've read the next chapter that the function $f : x \to \bigcup x$ which has $f(w) = a$ exists, i.e. a choice function exists.

In the same way, it can be shown that any finite family of non–empty sets has a choice function (because it follows from the Axioms of Pairing and Union show that every finite collection of sets is a set). But a problem arises when we have an *infinite* family x of non–empty sets: It is always possible to choose an element from each $w \in x$. What is not clear is that one can do this "in one go", i.e. simultaneously for all $w \in x$, via some function f.

□

I'm not sure if one can justify the Axiom of Choice via the iterative conception of sets. However, in 1938 Kurt Gödel proved the following result: If ZF (i.e. set theory *without* AC) is consistent, then so is ZFC (set theory with AC). Put differently, AC cannot introduce an inconsistency into set theory. If ZFC is inconsistent, then ZF was already inconsistent also.

ZFC 8: The Axiom of Replacement

The remaining two axioms — Replacement and Foundation — are very important for the subject of set theory itself, and for metamathematics. They play only a tiny role in ordinary mathematics however.

I do not see know to derive the Axiom Schema of Replacement from the iterative conception of sets. It is easy to derive it from the idea of *limitation of size*, however. By that we mean that a set is a collection that is not "too big". Here's how:

[15] We shall soon see — and you are no doubt aware already — that functions are sets. Remember, sets are *all* there is.

Suppose that $\varphi(x, y, p)$ is a first–order formula in the language of set theory which defines a "function". Here x, y play the role of independent and variable, and p the role of a fixed parameter. Informally, such a formula defines a "function" F when for any x there is at most one y such that $\varphi(x, y, p)$ holds. In that case we have a "function" F given by

$$y = F(x) \qquad \text{if and only if} \qquad \varphi(x, y, p) \text{ holds}$$

The Axiom of Relacement asserts that if F is such a "function" and if X is a set, then the direct image

$$F[X] := \{F(x) : x \in X\} = \{y : \exists x \in X \; \varphi(x, y, p)\}$$

is a set also. The use of quotation marks around the word *function* is essential here. We shall see in the next chapter how to interpret functions as sets. But it will follow from $ZFC\ 0-5$ that the direct image of a set under a function is a set — The Axiom of Replacement isn't necessary for that. The function F we use here may be "too big" to be a set, however. Nevertheless, even though F may be too big, its image $F[X]$ is surely no bigger than X. So if X is a set, then $F[X]$ can't be too big to be a set.

All this may be confusing at first. Fortunately we don't need to talk about the size of a "function" F to formulate the Axiom Schema of Replacement:

ZFC 8_φ:
Suppose that $\varphi(x, y, p_1, \ldots, p_n)$ is a first–order formula with free variables amongst x, y and $p := (p_1, \ldots, p_n)$. Suppose that for every x there is at most one y so that $\varphi(x, y, p)$. If X is a set, then the collection $\{y : \exists x \in X \; \varphi(x, y, p)\}$ is a set.

$$\forall p \left(\forall x \, \forall y \, \forall z \, (\varphi(x, y, p) \wedge \varphi(x, z, p) \to y = z) \;\to\; \forall X \exists Y \, \forall y \, (y \in Y \;\leftrightarrow\; \exists x \in X \; \varphi(x, y, p)) \right)$$

Observe that Replacement — like Separation — is an Axiom Schema. We have one axiom for every φ.

The Axiom Schema of Replacement was suggested independently by Abraham Fraenkel and Thoralf Skolem in 1922. It allows the creation of many, many new sets. Whereas the other axioms of ZFC allow one to form new sets that are at most a few levels higher than the sets from which they are created, the Axiom of Replacement allows one to form sets at very high levels. (E.g. it is possible that the sets X is at a "low level" and that the created set $F[X]$ at a very "high level".) In particular, it necessary to develop a theory of ordinal numbers and transfinite induction. As we shall see, the ordinals form the backbone that supports the entire set–theoretic universe, so this axiom is indispensable to the set theorist.

ZFC 9: The Axiom of Foundation

> *A well-known scientist (some say it was Bertrand Russell) once gave a public lecture on astronomy. He described how the earth orbits around the sun and how the sun, in turn, orbits around the center of a vast collection of stars called our galaxy. At the end of the lecture, a little old lady at the back of the room got up and said: "What you have told us is rubbish. The world is really a flat plate supported on the back of a giant tortoise." The scientist gave a superior smile before replying, "What is the tortoise standing on?" "You're very clever, young man, very clever," said the old lady. "But it's turtles all the way down!"*
>
> — Stephen Hawking, A Brief History of Time

In the set–theoretic universe, its sets all the way down. Every element of a set is a set. Every element of an element of a set is a set. But how far is *all the way down*? Not that far, as it turns out.

According to the iterative conception of sets, sets are made in stages, with each set being made up of elements that have been formed at a strictly earlier stage. We saw also that any non–empty collection of stages has a least element. Suppose now that x is a non–empty set. Let Γ be the collection of all stages associated with an element of x, i.e.

$$\Gamma = \{\gamma : \text{There is } x \in X \text{ such that } z \text{ is formed at stage } \gamma\}$$

The Γ is a non–empty collection of stages, so has a least element α. Let $y \in X$ be an element formed at stage α. If $z \in y$, then z is formed at a stage $\beta < \alpha$. Now $\beta \notin \Gamma$, because α is least, and hence $z \notin x$. It follows that if $z \in y$, then $z \notin x$, i.e. that $y \cap x = \varnothing$. We have thus shown (informally, from the iterative conception of sets), that for every non–empty set x there is $y \in x$ such that $y \cap x = \varnothing$.

We now take this to be an axiom:

ZFC 9:

$$\forall x \, (x \neq \varnothing \;\rightarrow\; \exists y \in x(y \cap x = \varnothing)$$

Exercise 1.4.11 Show using the Axiom of Foundation that "the turtles don't go very far down" in ZFC: There are is no sequence of sets $x_1, x_2, x_3, \ldots$ such that

$$\ldots x_{n+1} \in x_n \in \ldots x_3 \in x_2 \in x_1$$

It follows that every $\in$–decreasing chain

$$x_1 \ni \; x_2 \ni x_3 \ni \ldots$$

is of finite length. (Technically, we say that the $\in$–relation is *wellfounded*.)

□

The Axiom of Foundation is also referred to as the *Axiom of Regularity* in the literature. It was introduced by John von Neumann in 1925. While it is of scant use to the ordinary mathematician, it is of great import to the set theorist, as we shall see in a later chapter. In essence, together with Replacement, it allows us to formalize the iterative conception of set within ZFC itself, and build up the universe of sets from nothing, by iterating the operations of power set and union along the ordinals.

1.5 Sets and Classes

The version of set theory that we have developed (i.e. ZFC) has just one type of object, namely sets. Some set theories allow for another kind of object, however, namely *classes.* We will use this terminology only informally. Thus, informally, a class is a collection of the form

$$\mathbf{C} := \{x : \varphi(x, p)\}$$

i.e. $\mathbf{C}$ is the collection of all sets that have property φ. Here x, p are sets, with p playing the role of a parameter.

For example, the *universe of all sets* — denoted by $\mathbf{V}$ — is the class

$$\mathbf{V} := \{x : x = x\}$$

Some classes are “too big” to be sets: Russell’s paradox (together with Separation) shows that $\mathbf{V}$ is not a set. But *every set is a class*, for if y is a set, then the class determined by the formula $\varphi(x, y) \equiv x \in y$ is clearly y itself:

$$\{x : x \in y\} = y$$

A class that is not a set is called a *proper class.* Thus $\mathbf{V}$ is a proper class. Later, we shall see that the class **On** of all ordinals is a proper class also.

The theory NBG (von Neumann–Bernays–Gödel) is a commonly used set theory allows both classes and sets as objects. (This is a *conservative extension* of ZFC: Any theorem about sets is NBG is also a theorem of ZFC, and vice versa. Thus NBG says nothing about sets that ZFC doesn’t already know.)

Inside ZFC, however, classes are just a *way of speaking*: If

$$\mathbf{C} := \{x : \varphi(x, p)\}$$

is a class, then when we say $x \in \mathbf{C}$, we mean simply that the statement $\varphi(x, p)$ is true. Thus any mention of $\mathbf{C}$ can be avoided, simply by using its defining formula φ.

Exercise 1.5.1 Observe that every element of a class is a set. Show that a class is a set if and only if it is an element of some class.

□

In Exercise 1.4.7 it was shown that the intersection of a non–empty class is a set, i.e. that if $\mathbf{C} \neq \varnothing$, then $\bigcap \mathbf{C}$ is a set.

The Axiom Schema of Separation states essentially that a subclass of a set is a set:

$$\{x \in z : \varphi(x, p)\} = z \cap \{x : \varphi(x, p)\} = z \cap \mathbf{C}$$

The Axiom Schema of Replacement can be formulated as follows: If $\mathbf{F}$ is a class function[16], and X is a set, the direct image $\mathbf{F}[X]$ is a set.

[16] i.e. a "function" between classes. For example, the successor function $S(x) := x \cup \{x\}$ is a class function $S : \mathbf{V} \to \mathbf{V}$.

Chapter 2

Relations and Functions

In this chapter, we revise some elementary notions in mathematics, and show how they can be interpreted in the universe of sets. It is assumed that you've encountered the material in §2.1-2.5 before, so we move at speed. For the development below, the axioms $ZFC\ 0-5$ usually suffice. The Axiom of Choice (AC, ZFC 7) also occurs in a few exercises, but its use is always stated explicitly.

2.1 Ordered Pairs and Finite Cartesian Products

We can use the notion of an unordered pair to define the notion of an *ordered pair.* An ordered pair (x, y) is an object which, intuitively, contains x, y *in that order.* The essential property that we want ordered pairs to possess is the following:

$$(x,y) = (a,b) \qquad \text{if and only if} \qquad x = a \text{ and } y = b$$

In particular, $(x, y) \neq (y, x)$ (unless $x = y$). But by the Axiom of Extensionality

$$\{x, y\} = \{y, x\}$$

We can, however, *interpret* ordered pairs as sets, i.e. find sets which exhibit the correct behaviour. We follow the Polish mathematician Kazimierz Kuratowski, and define

$$\langle x, y\rangle := \{\{x\}, \{x, y\}\}$$

Exercise 2.1.1 Show that Kuratowski's definition of orded pair results in the correct behaviour, i.e. that $\langle x, y\rangle = \langle a, b\rangle$ if and only if $x = a$ and $y = b$. Also show that the following alternative definitions of ordered pair fail to exhibit the desired behaviour:

$$(a, b) := \{\{a\}, \{b\}\}, \qquad (a, b) := \{a, \{b\}\}$$

Finally, give one other definition of ordered pair that *does* exhibit the correct behaviour.

□

Observe that the set $\langle x, y\rangle$ is a concrete realization of the abstract idea of an ordered pair (x, y). This is not to say that (x, y) *is* $\langle x, y\rangle$, but merely that the set $\langle x, y\rangle$ *behaves* as an ordered pair should.

Once we have defined ordered pairs as sets, we are able to find interpretations of ordered triples and quadruples,

$$\langle x, y, z\rangle := \langle\langle x, y\rangle, z\rangle, \qquad \langle x, y, z, w\rangle := \langle\langle x, y, z\rangle, w\rangle,$$

and, in general[1],

$$\langle x_1, \dots, x_{n+1}\rangle := \langle\langle x_1, \dots, x_n\rangle, x_{n+1}\rangle$$

Given two sets X, Y, we may define the Cartesian product $X \times Y$ of X and Y as follows:

$$X \times Y = \{z : \exists x \in X\ \exists y \in y\ (z = \langle x, y\rangle)\}$$

We then define

$$X \times Y \times Z := (X \times Y) \times Z$$

etc. Observe that in that case

$$X \times Y \times Z = \{\langle x, y, z\rangle : x \in X, y \in Y, z \in Z\}$$

Similarly, we define $X^2 := X \times X$, $X^3 := X \times X \times X$, etc.

We have not yet shown that Cartesian products are sets. That is the objective of the next exercise:

Exercise 2.1.2 Let X, Y be sets.

(a) Write down a first–order formula *pair*(z, x, y) which states that $z = \{x, y\}$.

(b) Write down a first–order formula *ordered_pair*(z, x, y) in the langage of set theory which states that $z = \langle x, y\rangle$.

(c) Show that if $x \in X$ and $y \in Y$, then $\langle x, y\rangle \in \mathcal{P}(\mathcal{P}(X \cup Y))$.

(d) Show that $X \times Y$ is a set.

□

Here are some easy exercises about the arithmetic of Cartesian products:

Exercise 2.1.3 (a) $(X \cup Y) \times Z = (X \times Z) \cup (Y \times Z)$, $(X \cap Y) \times Z = (X \times Z) \cap (Y \times Z)$, $(X - Y) \times Z = (X \times Z) - (Y \times Z)$,

(b) $(A \times B) \cap (X \times Y) = (A \cap X) \times (B \cap Y)$.

(c) $X \times Y = \varnothing$ if and only if $X = \varnothing$ or $Y = \varnothing$.

□

[1] We are not using natural numbers in this definition, but giving a schema for defining ordered tuples of arbitrary length. If you want to define an ordered 100–tuple, this requires a formula containing 100 variables. The concept/number 100 is not required. However, in order to define all finite ordered tuples simultaneously, using induction, we need the concept of natural number. This will have to wait until a later chapter.

2.2 Relations

In mathematics, one studies many *relations* between various objects x, y. For example:

- $x < y$ where x, y are numbers.
- $x \in y$, where x, y are sets.
- $x = y$, where x, y are natural numbers/real numbers/sets/matrices/...
- $x|y$ (x divides y), where x, y are integers, or polynomials.
- x *is a root of* y, where x is a real number, and y a real function.

If R stands for some relation, we often write xRy to mean that x, y satisfy that relation. Relations need not be *binary*, i.e. they need not be between just two objects. Furthermore, the objects need not be of the same type[2]. For example:

- x, y, z are related if and only if x is congruent to y modulo z, where x, y, z are natural numbers.
- x, y, z are related if and only if $x(y) = z$, where x is a real function, and y, z are real numbers.
- x, y, z are related if and only if the dot product of x and y is z, where x, y are two n–dimensional vectors, and z is a real number.
- x, y, z, w are related if and only if the point (x, y) lies on a circle of radius z and centre w.

Observe that the order of $x, y, z, \ldots$ may matter. If the relation is *strict inequality* $<$, then x is related to y if and only if $x < y$. This does not mean y is also related to x.

Given some relation, we can interpret it as a set R as follows: Let R be the set of all ordered pairs $\langle x, y\rangle$ with the property that x is related to y. Thus

Definition 2.2.1 A *binary relation* R is a set of ordered pairs. When $\langle x, y\rangle \in R$, we may indicate this by writing xRy.
An n–ary relation R is a set of ordered n–tuples. When $\langle x_1, \ldots, x_n\rangle \in R$, we may indicate this by writing $R(x_1, \ldots, x_n)$.

□

For example, the relation of strict inequality on the natural numbers[3] $\mathbb{N} := \{0, 1, 2, 3, \ldots\}$ is the set $<$ defined by

$$<:= \Big\{\langle 0,1\rangle, \langle 0,2\rangle, \langle 0,3\rangle, \langle 0,4\rangle, \ldots, \langle 1,2\rangle, \langle 1,3\rangle, \langle 1,4\rangle, \ldots, \langle 2,3\rangle, \langle 2,4\rangle, \ldots, \langle 3,4\rangle, \ldots\Big\}$$

Definition 2.2.2 Suppose that R is a binary relation.

[2] Informally speaking, that is. In the end of course, they will be of the same type, since everything is a set.

[3] We will define the natural numbers later in this chapter.

(a) The *domain* and *range* of R are defined by

$$\mathrm{dom}(R) := \{x : \exists y\ (xRy)\} \qquad \mathrm{ran}(R) := \{y : \exists x\ (xRy)\}$$

(b) If A is a set, then the *image* of A under R is defined by

$$R[A] := \{y \in \mathrm{ran}(R) : \exists x \in A\ (xRy)\}$$

(c) If B is a set, then the *inverse image* of B under R is defined by

$$R^{-1}[B] := \{x \in \mathrm{dom}(R) : \exists y \in B\ (xRy)\}$$

(d) If A is a set, then the *restriction* of R to A is the binary relation $R \restriction A$ defined by

$$R \restriction A := \{\langle x, y\rangle \in R : x \in A\}$$

□

If X, Y are sets, then any subset $R \subseteq X \times Y$ is a relation, with $\mathrm{dom}(R) \subseteq X$, and $\mathrm{ran}(R) \subseteq Y$. In that case, we say that R is a relation *from* X *to* Y. If $R \subseteq X \times X$, we say that R is a relation *on* X.

For any set X there is a *identity relation* Id_X on X defined by

$$\mathrm{Id}_X := \{\langle x, x\rangle : x \in X\}$$

so that $x\ \mathrm{Id}_X\ y$ if and only if $x, y \in X$ and $x = y$. (Thus Id_X is just the relation of equality, restricted to the set X.)

Exercise 2.2.3 Let R be a binary relation.

(a) Show that if $\langle x, y\rangle \in R$, then $x, y \in \bigcup\bigcup R$. Conclude that that $\mathrm{dom}(R)$ and $\mathrm{ran}(R)$ are sets.

(b) Show that $R \subseteq \mathrm{dom}(R) \times \mathrm{ran}(R)$.

(c) Define the *inverse* R^{-1} of R by

$$\langle x, y\rangle \in R^{-1} \qquad \text{iff} \qquad xRy$$

Show that R^{-1} is a set, and thus a binary relation.

(d) If S is another binary relation, define the *composition* $S \circ R$ of R with S by

$$\langle x, z\rangle \in S \circ R \qquad \text{iff} \qquad \exists y\ (xRy \wedge ySz)$$

Show that $S \circ R$ is a set, and thus a binary relation.

(e) Show that composition of binary relations is *associative*:

$$T \circ (S \circ R) = (T \circ S) \circ R$$

(f) Consider the relation $\leq$ on the set of natural numbers. What is $\leq^{-1}$? Show that $\leq \circ \leq = \leq$.

(g) Show that if $\mathrm{dom}(R) = X, \mathrm{ran}(R) \subseteq Y$, then $\mathrm{Id}_Y \circ R = R = R \circ \mathrm{Id}_X$.

□

Observe that $xS \circ Rz$ if and only if there is y such that $xRySz$ — the order of R, S is reversed! (This is why some texts write $R \circ S$ where we write $S \circ R$. We have reasons for preferring the current version, however, as will become apparent in the next section.)

Definition 2.2.4 A binary relation R on a set X is

(a) *Reflexive* if $\forall x \in X\ (xRx)$.

(b) *Symmetric* if $\forall x, y \in X\ (xRy \to yRx)$.

(c) *Asymmetric* if $\forall x, y \in X\ (xRy \to \neg(yRx))$.

(d) *Antisymmetric* if $\forall x, y \in X\ (xRy \wedge yRx \to x = y)$.

(e) *Transitive* if $\forall x, y, z \in X\ (xRy \wedge yRz \to xRz)$.

(f) *Connected* if $\forall x, y \in X\ (xRy \vee x = y \vee yRx)$.

□

Exercise 2.2.5 Suppose that R is a binary relation on a set X.

(a) Show that R is reflexive if and only if $\mathrm{Id}_X \subseteq R$.

(b) Show that R is symmetric if and only if $R^{-1} \subseteq R$.

(c) Show that R is asymmetric if and only if $R \cap R^{-1} = \varnothing$.

(d) Show that R is transitive if and only if $R \circ R \subseteq R$.

(e) Show that R is connected if and only if $R \cup R^{-1} \cup \mathrm{Id}_X = X \times X$.

□

2.3 Functions

Informally, a function f can be thought to induce a binary relation between elements x, y, namely: x and y are related if $f(x) = y$. But not every binary relation is a function, for functions are *single–valued*, i.e. if $f(x) = y$ and $f(x) = z$, then we must have $y = z$. This leads to the following definition:

Definition 2.3.1 A *function* (or *map*) f is a binary relation with the property that if $\langle x, y\rangle \in f$ and $\langle x, z\rangle \in f$, then $y = z$.
We may write $f(x) = y$ instead of $\langle x, y\rangle \in f$ or $x\ f\ y$.

□

When f is a function with $\mathrm{dom}(f) = X$ and $\mathrm{ran}(f) \subseteq Y$, we indicate this by writing $f : X \to Y$. We may also write $x \mapsto y$ to mean that $f(x) = y$ (when f is clear from context).

If X, Y are sets, then the collection of all functions $f : X \to Y$ is denoted by Y^X. (Some authors use ${}^X Y$.)

Exercise 2.3.2 Show that if X, Y are sets, then X^Y is a set.

□

Observe that the identity relation Id_X on a set X is a *function*, with $\mathrm{Id}_X : X \to X : x \mapsto x$. Obviously, it will be referred to as the *identity function* as well.

Since a function is a binary relation, the notions of domain, range, inverse, image, inverse image, restriction and composition make sense. In particular any function f will possess an inverse relation f^{-1}. However, f^{-1} need not be a function.

Definition 2.3.3 Let $f : X \to Y$ be a function.

(a) If the binary relation f^{-1} is a function $f : Y \to X$, i.e. if f^{-1} is a function and $\mathrm{dom}(f^{-1}) = Y$, then we say that f is *invertible*.

(b) We say $f : X \to Y$ is *injective* (or *one-to-one*) if $\forall x, x' \in X\ f(x) = f(x') \to x = x')$. We may write $f : X \rightarrowtail Y$ (or $f : X \hookrightarrow Y$) to assert that f is an injection.

(c) We say that $f : X \to Y$ is *surjective* (or *onto*) if $Y = \mathrm{ran}(f)$, i.e. if $\forall y \in Y\ \exists x \in X\ (f(x) = y)$. We may write $f : X \twoheadrightarrow Y$ to assert that f is a surjection.

(d) We say that $f : X \to Y$ is a *bijective* if f is *both* injective and surjective. We may write $f : X \rightarrowtail\!\!\!\!\rightarrow Y$ to assert that f is a bijection.

□

Exercise 2.3.4 Let $f : X \to Y$ be a function.

(a) Show that a composition of in/sur/bijections is an in/sur/bijection.

(b) Show that the relation f^{-1} is a function if and only if f is injective.

(c) Show that if f is invertible, then $f \circ f^{-1} = \mathrm{Id}_Y$, and $f^{-1} \circ f = \mathrm{Id}_X$.

(d) Say that a function $g : Y \to X$ is a *left inverse* of f if $g \circ f = \mathrm{Id}_X$. Show that f has a left inverse if and only if it is an injection.

(e) Say that a function $h : Y \to X$ is a *right inverse* of f if and only if $f \circ h = \mathrm{Id}_Y$. Using AC, show that a function has a right inverse if and only if it is a surjection.

(f) Show that if f has both a left inverse g and a right inverse h, then f is invertible, and $g = h = f^{-1}$.

(g) Show that f is invertible if and only if it is a bijection.

□

Suppose that X is a sets, and that $F : I \to X : i \mapsto x_i$ is a surjection. Then

$$X = \{F(i) : i \in I\} = \{x_i : i \in I\}$$

The function F is called an *indexing* of the set X. The set I is called the *index set*, each $i \in I$ is called an *index*, and $F(i) = x_i$ is called the i^{th} *term*.

Note that any set can be indexed by some set. For example, a set X can be indexed by itself: Let $I = X$, and let $x_i = i$. Later on, however, we shall often use *ordinals*, a kind of transfinite number, to index sets.

If $X = \{x_i : i \in I\}$, we will write

$$\bigcup X = \bigcup_{i \in I} x_i \qquad \bigcap X = \bigcap_{i \in I} x_i$$

etc.

Exercise 2.3.5 Let $f : X \to Y$ be a function. Show that the inverse image $f^{-1}[\cdot]$ is a function

$$f^{-1}[\cdot] : \mathcal{P}(Y) \to \mathcal{P}(X) : B \mapsto f^{-1}[B] := \{x \in X : f(x) \in B\}$$

which preserves the basic set–theoretic operations $\bigcup, \bigcap, -$, (e.g. show that $f^{-1}[\bigcup_{i \in I} Y_i] = \bigcup_{i \in I} f^{-1}[Y_i]$ for any family $Y_i \in \mathcal{P}(Y)$.)

□

Associated with any product $X \times Y$ are associated *projections* π_X, π_Y defined by

$$\pi_X : X \times Y \to X : \langle x, y\rangle \mapsto x \qquad \pi_Y : X \times Y \to Y : \langle x, y\rangle \mapsto y$$

Exercise 2.3.6 Suppose that $f_X : A \to X$, $f_Y : A \to Y$ are functions with common domain A. Show that there is a unique function $f : A \to X \times Y$ such that $f_X = \pi_X \circ f$ and $f_Y = \pi_Y \circ f$. [Hint: Let $f \subseteq A \times (X \times Y)$ consist of all ordered pairs $\langle a, \langle f_X(a), f_Y(a)\rangle\rangle$ where $a \in A$.

□

2.4 Equivalence Relations

Recall that any set X carries an *identity relation* Id_X defined by

$$\mathrm{Id}_X := \{\langle x, x\rangle : x \in X\}$$

which is just the relation of equality, restricted to the set X.

The relation of equality is a class relation which satsifies following properties:

- $=$ is reflexive: $x = x$.
- $=$ is symmetric. If $x = y$, then $y = x$.
- $=$ is transitive: If $x = y$ and $y = z$, then $x = z$.

A relation which satisfies these same properties is called an *equivalence relation*. To be precise:

Definition 2.4.1 A binary relation R on a set X is an equivalence relation if and only if

(i) R is *reflexive*, i.e. $\forall x \in X \ (xRx)$.

(ii) R is *symmetric*, i.e. $\forall x, y \in X \ (xRy \ \rightarrow \ yRx)$.

(iii) R is *transitive*, i.e. $\forall x, y, z \in X \ (xRy \ \wedge \ yRz \ \rightarrow \ xRz)$.

□

It is easy to see that if n is a non–zero integer, then the relation $\equiv_n$ of *equivalence modulo*[4] n is an equivalence relation on the set $\mathbb{Z}$ of integers.

An equivalence relation R can be thought of as a generalization of the identity relation: If xRy, then x, y are in a certain sense identified, i.e. regarded as "equal" for the purpose at hand. When we collect together all the elements which are made "equal" we get equivalence classes:

Definition 2.4.2 If R is an equivalence relation on a set X and $x \in X$, then the *equivalence class* of x *modulo* R is the set defined by:

$$[x]_R := \{y \in X : xRy\}$$

□

Some authors may write x/R instead of $[x]_R$.

The following device yields a fruitful way of looking at equivalence relations:

Definition 2.4.3 A family $\mathcal{P}$ of *non–empty* sets is called a *partition* of a set X if and only if:

(i) $\bigcup\{P : P \in \mathcal{P}\} = X$.

(ii) Any two distinct members of $\mathcal{P}$ are disjoint, i.e. $P, Q \in \mathcal{P}$ and $P \neq Q$ implies $P \cap Q = \varnothing$.

□

The members $P \in \mathcal{P}$ are sometimes called the *blocks* of the partition. Observe that (i) of the above definition states that each $x \in X$ belongs to at least one block, whereas (ii) states that no $x \in X$ belongs to more than one block. Thus each $x \in X$ belongs to exactly one block.

Exercise 2.4.4 (a) *Every equivalence relation on X induces a partition of X:* Let R be an equivalence relation on a set X. Define a family of sets $\mathcal{P}_R$ by

$$\mathcal{P}_R := \{[x]_R : x \in X\}$$

Show that $\mathcal{P}_R$ is a partition of X.

[4] Recall that $a \equiv_n b$ if and only if $b - a$ is divisible by n.

(b) *Every partition of a set X induces an equivalence relation on X:* Let $\mathcal{P}$ be a partition of X. Define a binary relation $R_{\mathcal{P}}$ on X by

$$x \; R_{\mathcal{P}} \; y \qquad \text{if and only if} \qquad \exists P \in \mathcal{P} \; (x \in P \wedge y \in P)$$

(i.e. $x \; R_{\mathcal{P}} \; y$ if and only if x, y belong to the same block of $\mathcal{P}$.) Show that $R_{\mathcal{P}}$ is an equivalence relation. Further show that the blocks of $\mathcal{P}$ are precisely the equivalence classes of $R_{\mathcal{P}}$, i.e. that $[x]_{R_{\mathcal{P}}} = P$ if and only if $x \in P$.

(c) Show that the constructions in (a), (b) above are inverses of each other, in the following sense: Starting from an equivalence relation R, the construction in (a) yields a partition $\mathcal{P}_R$. If we then apply the construction in (b) to the partition $\mathcal{P}_R$, we get an equivalence relation $R_{\mathcal{P}_R}$. This relation is precisely the original equivalence relation R, i.e. $R_{\mathcal{P}_R} = R$. Similarly, $\mathcal{P}_{R_{\mathcal{P}}} = \mathcal{P}$.

□

If R is an equivalence relation on a set X, then the set of equivalence classes — which we above denoted by $\mathcal{P}_R$ — is usually denoted by X/R. The equivalence class of an element x — above denoted by $[x]_R$ — is often denoted by x/R.

Here is another reason why equivalence relations are ubiquitous in mathematics:

Exercise 2.4.5 (a) *Every function induces an equivalence relation on its domain:* Suppose that $f : X \to Y$ is a function. Define a binary relation R on X by

$$xRy \qquad \text{iff} \qquad f(x) = f(y)$$

The relation R is called the *kernel* of f, and denoted $R = \ker f$. Show that $\ker f$ is an equivalence relation.

(b) Assuming the Axiom of Choice, *every equivalence relation on a set X is induced by a function with domain X:* Show that if R is an equivalence relation, then there is a function f with domain X such that $R = \ker f$.

□

The following exercise gives two equivalents of AC:

Exercise 2.4.6 Show that the following are equivalent:

(a) AC

(b) For every partition $\mathcal{P}$ there is a set C which has precisely one element in common with each block $P \in \mathcal{P}$.

(c) Every surjection has a right inverse.

□

2.5 Order Relations

Definition 2.5.1 (a) A binary relation $\leq$ on a set X is said to be a *partial order* relation if $\leq$ is:

(i) reflexive, i.e. $x \leq x$,

(ii) antisymmetric, i.e. $x \leq y$ and $y \leq x$ imply $x = y$, and

(iii) transitive, i.e. $x \leq y$ and $y \leq z$ imply $x \leq z$.

(b) A binary relation $<$ on a set X is said to be a *strict partial order* relation if $\leq$ *is*

(i) asymmetric, i.e. $x < y$ implies $y \not< x$.

(ii) transitive, i.e. $x < y$ and $y < z$ imply $x < z$.

(c) A *total ordering* is a connected partial ordering. A *strict total ordering* is a connected strict partial ordering. A (strictly) totally ordered set is also called a *chain*.

□

A *partially ordered set* is a pair $(X, \leq)$ where $\leq$ is a partial ordering on X.

Observe that Id_X is a partial ordering on X, and that $\varnothing$ is a strict partial ordering on X.

Given a partial ordering $\leq$, we can define a strict partial ordering $<$ by

$$x < y \qquad \text{if and only if} \qquad x \leq y \wedge x \neq y$$

i.e. $< := \leq - \mathrm{Id}_X$.

Conversely, given a strict partial ordering $<$, we can define a partial ordering $\leq$ by

$$x \leq y \qquad \text{if and only if} \qquad x < y \vee x = y$$

i.e. $\leq := < \cup \mathrm{Id}_X$.

Moreover, $<$ is total if and only if $\leq$ is total.

Definition 2.5.2 Let $\leq$ be a partial ordering on X.

(a) Let $Y \subseteq X$. An element $y_0 \in Y$ is said to be a $\leq$–*minimal element* of Y if there is no $y \in Y$ such that $y < y_0$.
We will refer to y_0 simply as a minimal element of Y if $\leq$ is clear from context.

(b) Let $Y \subseteq X$. An element $y_0 \in Y$ is said to be the $\leq$–*minimum* element of Y if and only if $\forall y \in Y\ (y_0 \leq y)$.
We will refer to y_0 simply as the minimum element of Y, or the least element of Y, if $\leq$ is clear from context.

(c) The partial ordering $\leq$ is said to be *well–founded* if every non–empty subset of X has an $\leq$–minimal element.

(d) A total ordering $\leq$ is called a *well–ordering* if every non–empty subset of X has a $\leq$–minimum element.

□

Exercise 2.5.3 Let $(X, \leq)$ be a partial ordering, and let $Y \subseteq X)$.

(a) Every minimum element of Y is a minimal element. The converse may be false.

(b) A subset of a partial ordering can have at most one minimum element.

(c) If $(X, \leq)$ is a chain, then every minimal element of Y is also a minimum element of Y.

□

Definition 2.5.4 Let $(A, \leq)$ and $(B, \preceq)$ be partially ordered sets.

(a) A function $f : A \to B$ is *order–preserving* if and only if

$$x \leq y \;\to\; f(x) \preceq f(y)$$

(b) $(A, \leq)$ and $(B, \preceq)$ are said to be *order–isomorphic* if there is a bijection $f : A \to B$ with the property that both f, f^{-1} are order preserving. Such an f is called an *order–isomorphism.*

□

Exercise 2.5.5 (a) Let $(A, \leq)$ and $(B, \preceq)$ be partially ordered sets. Show that a bijection $f : A \to B$ is an order–isomorphism if and only

$$x \leq y \;\leftrightarrow\; f(x) \leq f(y)$$

(b) Give an example to show that not every order–preserving bijection is an order–isomorphism.

(c) Show that $(A, \leq)$ and $(B, \preceq)$ are chains, then any order–preserving bijection is an order–isomorphism.

(d) Show that if X is a set and $\mathcal{X}$ is a family of subsets of X, then the inclusion relation $\subseteq$ is a partial ordering on $\mathcal{X}$, i.e. $(\mathcal{X}, \subseteq)$ is a partial ordering.

(e) Let $(X, \leq)$ be a partial ordered set. For $x \in X$, define

$$\downarrow x := \{y \in X : y \leq x\} \qquad \text{and} \qquad \mathcal{X} := \{\downarrow x : x \in X\}$$

Show that $(X, \leq)$ and $(\mathcal{X}, \subseteq)$ are order–isomorphic.
(Thus, in some sense, $\subseteq$ is the *only* partial order relation that exists, up to isomorphism.)

□

Chapter 3

Natural Numbers

God created the natural numbers; everything else is the work of man.
— attributed to Leopold Kronecker

It seems that Kronecker may have almost literally believed this. When Lindemann proved in 1882 that π is trancendental (i.e. not the root of a polynomial with rational coefficients), Kronecker opined that the proof was indeed beautiful, but since irrational numbers do not exist, Lindemann's proof had no meaning.
Certainly, Kronecker was the first and most vociferous critic of Cantor's transfinite sets [1].

In 1889, Giuseppe Peano listed five axioms with the intention of providing a rigorous foundation for the natural numbers. The *Peano Postulates* are:

I. 0 is a natural number.

II. Every natural number has a successor $s(n)$ which is also a natural number.
(The intention is that $s(n) = n + 1$.)

III. 0 is not the successor of any natural number.

IV. If two natural numbers have the same successor, then they are identical, i.e. if $s(n) = s(m)$, then $n = m$.

V. If X is a set of natural numbers such that:

   (i) $0 \in X$,
   (ii) If $n \in X$, then $s(n) \in X$,

   then every natural number belongs to X.
   (This is the *Principle of Mathematical Induction.*)

The informal idea is that the set of natural numbers is the set

$$\{0,\ s(0),\ s(s(0)),\ s(s(s(0))),\ \ldots\}$$

so that every natural number is obtained from 0 by finitely many applications of the successor function $s(\cdot)$. The point of this chapter is to create the natural numbers: We shall construct an object which satisfies the Peano Postulates inside the set–theoretic universe.

[1] Doron Zeilberger is a well–known modern mathematician who holds similar opinions and publishes them regularly. They're worth reading. Go have a look on the www.

In this chapter, we work mainly with the theory ZF$^-$, which is ZFC *without* the Axioms of Choice and Foundation. AC will be used occasionally, but its use will in all cases be stated expplicitly.

3.1 The Set ω of Natural Numbers

Let us briefly recall some definitions made when we introduced the Axiom of Infinity: We defined the *successor function* $S(\cdot)$ on sets by

$$S(x) := x \cup \{x\}$$

(Some authors write x^+ instead of $S(x)$.) We defined a set X to be *inductive* if

(i) $\varnothing \in X$, and

(ii) X is closed under $S(\cdot)$: If $x \in X$, then $S(x) \in X$.

The Axiom of Infinity is the statement

There exists an inductive set

Now let $\mathbf{I}$ be the class of all inductive sets:

$$\mathbf{I} := \{x : \mathrm{Ind}(x)\} \qquad \text{where} \quad \mathrm{Ind}(x) \equiv \varnothing \in x \ \wedge \ \forall y\ (y \in x \to S(y) \in x)$$

By the Axiom of Infinity, $\mathbf{I}$ is non–empty. It follows from Exercise 1.4.7 that $\bigcap \mathbf{I}$ is a set. (Indeed, if A is any inductive set, then $\bigcap \mathbf{I} = \{x \in A : \mathrm{Ind}(x)\}$ is a set by Separation.)

We will give this set a name:

$$\omega := \bigcap \mathbf{I}$$

and call ω the *set of natural numbers.* Another commonly used symbol for the set of natural numbers is $\mathbb{N}$.

It is easy to see that ω is itself an inductive set. It follows that ω is the *smallest* inductive set: If x is any inductive set, then $\omega \subseteq x$.

Exercise 3.1.1 (a) Show that the intersection of a non–empty class of inductive sets is an inductive set.

(b) Conclude that ω is the smallest inductive set: If x is any inductive set, then $\omega \subseteq x$.

□

Since $\varnothing \in \omega$ and ω is closed under $S(\cdot)$, we see that ω must have the following elements (which we simultaneously provide with names):

$$\begin{aligned}
0 &:= \varnothing \\
1 &:= S(0) = \{0\} \\
2 &:= S(1) = \{0, 1\} \\
3 &:= S(2) = \{0, 1, 2\} \\
4 &:= S(3) = \{0, 1, 2, 3\} \\
&\ \vdots \\
n &:= \{0, 1, 2, \ldots, n-1\} \\
&\ \vdots
\end{aligned}$$

Observe that the set n has (informal) n elements.

If $n \in \omega$, we write $n+1 := S(n)$ (i.e. $n+1$ is another notation for $S(n) := n \cup \{n\}$).

The following result is easy:

Theorem 3.1.2 (Principle of Mathematical Induction) *Suppose that $X \subseteq \omega$ has the properties that*

(i) $0 \in X$.

(ii) If $n \in X$, then $n+1 \in X$.

Then $X = \omega$.

□

Exercise 3.1.3 Prove the Principle of Mathematical Induction.

□

Let us investigate the structure of the set ω in greater detail. We want to verify that ω, S satisfy the Peano Postulates. We won't take the simplest route — things are about to get quite technical — but the route we follow is good preparation for the general study of the ordinals. In particular, we will also study the order relation on the natural numbers.

3.2 $(\omega, \in)$ is a Transitive Well–Ordered Set

.

Above, we saw that $1 = \{0\}, 2 = \{0,1\}, 3 = \{0,1,2\} \ldots$ etc. Thus, intuitively, $m < n$ if and only if $m \in n$. In other words, it seems that the $\in$–relation acts as a strict total ordering $<$ on the set of natural numbers. Of course, the set of natural numbers yields the primary example of a *well–ordered set*[2]. It is the fact that the natural numbers are well–ordered that makes possible proofs by induction and definitions by recursion, so the importance well-ordering is hard to overstate.

In this section, we make precise and verify the above assertions for our interpretation ω of the set of natural numbers. We begin with some definitions:

Definition 3.2.1 (a) A set X is *transitive* if each element of X is also a subset of X:

$$\text{Trans}(X) \equiv \forall x \in X (x \subseteq X)$$

(This means that if $y \in x$ and $x \in X$, then also $y \in X$.)

(b) An element $y \in x$ is said to be an $\in$–*minimal* element of x if there is no $z \in y$ is such that $z \in x$. Put differently, $y \in x$ is an $\in$–minimal element of x if $y \cap x = \varnothing$.

(c) A set x is *well–founded*[3] if every non–empty subset $z \subseteq x$ has a $\in$–minimal element.

[2] Recall that a total ordering $(X, <)$ is a well–ordering if every non–empty subset of X has a least element. We will investigate well–ordered sets in more detail in the next chapter.

[3] By the Axiom of Foundation, *every* non–empty set has an $\in$–minimal element, i.e. every set is well–founded. But we do not want to use this axiom here if we can avoid it.

□

Lemma 3.2.2 *(a) If x is transitive, then so is $S(x)$.*

(b) If x is well–founded, then $x \notin x$.

(c) If x is transitive and well–founded, then so is $S(x)$.

(d) If x, y are sets such that (i): x is transitive and well–founded, and (ii): $S(x) = S(y)$, then $x = y$.

Proof: (a) If $y \in S(x) = x \cup \{x\}$, then either $y \in x$ or $y = x$. Since x is transitive, $y \subseteq x$ in either case, and thus $y \subseteq S(x)$ also.

(b) $\{x\}$ is a non–empty subset of x, and thus has a $\in$–minimal element, i.e. there is $y \in \{x\}$ so that $y \cap \{x\} = \varnothing$. But then $x = y$, so $x \cap \{x\} = \varnothing$, contradicting the fact that $x \in x \cap \{x\}$.

(c) We have already seen that $S(x)$ is transitive if x is. Now suppose further that x is well–founded, and that $y \subseteq S(x)$ is non–empty. We distinguish two cases:
Case (i): If $y \cap x$ is non–empty, then since x is well–founded, $y \cap x$ has an $\in$–minimal element z, i.e. there there is $z \in y \cap x$ such that $z \cap y \cap x = \varnothing$. We claim that z is an $\in$–minimal element of y also. For suppose that there is $w \in z \cap y$. Then certainly $w \in z$. But as $z \in x$ and x is transitive, we must have $w \in x$ as well. But then $w \in z \cap y \cap x$ — contradiction, because $z \cap y \cap x = \varnothing$.
Case (ii): If $y \cap x = \varnothing$, then $y = \{x\}$. But since we know that $x \notin x$, we see that x is a $\in$–minimal element of y.

(d) Suppose $x \neq y$. Then $x \in y$ and $y \in x$. Since x is transitive, $x \in x$. But since x is well–founded, $x \notin x$ —contradiction.

$\dashv$

Lemma 3.2.3 *(a) If X is inductive, then $\{x \in X : x \subseteq X\}$ is inductive also.
Hence ω is transitive.*

*(b) If X is inductive, then $\{x \in X : x$ is transitive and well-founded$\}$ is inductive also.
Hence every element $n \in \omega$ is transitive and well-founded.*

(c) ω is well–founded.

*(d) If X is inductive, then $\{x \in X : x = \varnothing \ \vee \ \exists y \in X\ (x = S(y))\}$ is inductive also.
Hence every non–0 element of ω is the successor of some element of ω.*

Proof: (a) Let $Y := \{x \in X : x \subseteq X\}$. Then $\varnothing \in Y$. Now suppose that $x \in Y$. Then $x \in X$, so $S(x) \in X$, because X is inductive. Further, $x \subseteq X$, and hence $S(x) \subseteq X$. Since $S(x) \in X$ and $S(x) \subseteq X$, we see that $S(x) \in Y$. Hence Y is inductive.
It follows that $\{n \in \omega : n \subseteq \omega\}$ is an inductive subset of ω. Since ω is the smallest inductive set, we see that $n \subseteq \omega$ holds for every $n \in \omega$. Thus ω is transitive.

(b) Let $Y = \{x \in X : x$ is transitive and well–founded$\}$. Certainly $\varnothing \in Y$. Now if $x \in Y$, then $S(x) \in X$ (because X is inductive). Moreover, $S(x)$ is transitive and well–founded (by Lemma 3.2.2). Hence $S(x) \in Y$, and thus Y is inductive.

It follows that $\{n \in \omega : n \text{ is transitive and well–founded}\}$ is an inductive subset of ω, and thus that it is equal to ω.

(c) Let $X \subseteq \omega$ be a non–empty set, and let $n \in X$. If $n \cap X = \varnothing$, then n is an $\in$–minimal element of X. Else $n \cap X$ is a non–empty subset of the well-founded set n, and therefore $n \cap X$ has an $\in$–minimal element m. We claim that m is an $\in$–minimal element of X also. Indeed, since $m \in n$ and n is transitive, we have $m \subseteq n$, and hence $m \cap X = m \cap n \cap X = \varnothing$.

(d) Let $Y := \{x \in X : x = \varnothing \ \vee \ \exists y \in X \ (x = S(y))\}$, so that a non–empty element of X belongs to Y if and only if it is the successor of some element of X. Then clearly $\varnothing \in Y$. Moreover, if $x \in Y$, then clearly $S(x)$ is the successor of some element of X, and hence $S(x) \in Y$ also.
As above, we see that $\{n \in \omega : n = 0 \vee \exists m \in \omega \ (n = S(m))\}$ is an inductive subset of ω, and hence is equal to ω.

$\dashv$

We can now see that ω, S do indeed satisfy Peano's axioms:

I. $0 \in \omega$, because ω is inductive.

II. If $n \in \omega$, then $S(n) \in \omega$, because ω is inductive.

III. $0 \neq S(n)$ for any n, because 0 is the empty set, whereas $n \in S(n)$.

IV. If $S(n) = S(m)$, then $n = m$, by Lemmas 3.2.2(d) and 3.2.3(b).

V. The Principle of Mathematical Induction holds, by Theorem 3.1.2.

Now let's investigate the order relation on the set of natural numbers. Define a binary relation $<$ on ω as follows:

$$m < n \qquad \text{if and only if} \qquad m \in n$$

As usual, "$n \leq m$" means "$n = m \ \vee \ n < m$".
Observe that $0 < 1 < 2 < 3 < \dots$, because $0 \in 1 \in 2 \in 3 \in \dots$.

Theorem 3.2.4 $(\omega, \in)$ *is a transitive strict well–ordered set.*

□

We already know that ω is a transitive well–founded set, and that the same is true for any $n \in \omega$. The proofs of the remaining assertions of this theorem are given in the following exercise:

Exercise 3.2.5 (a) Show that $0 \leq n$ for all $n \in \omega$.
[Hint: Let $X := \{n \in \omega : 0 \leq n\}$ and show that X is an inductive subset of ω.]

(b) Show that for all $m, n \in \omega$ we have $m < n + 1$ if and only if $m \leq n$.

(c) Show that for all $m, n \in \omega$, we have $m < n$ if and only if $m + 1 \leq n$.
[Hint: Let $X := \{n \in \omega : \forall m \in \omega \ (m < n \ \to \ m + 1 \leq n)\}$.]

(d) Show that the relation $<$ is *transitive*: If $k < m$ and $m < n$, then $k < m$.

(e) Show that the relation $<$ is *asymmetric*: If $n < m$, then $m \not< n$.

(f) Show that the relation $<$ is *connected*: For any n, m, either $n < m$ or $n = m$ or $m < n$. [Hint: Show, with the aid of (a)–(c) that the set $\{n \in \omega : \forall m \in \omega\ (n < m \ \vee\ n = m \ \vee\ m < n\}$ is inductive.]

(g) Show that if $X \subseteq \omega$ is non–empty, then X has a $<$–least element.

□

We thus see that ω is well–ordered by the $\in$–relation. Furthermore, if $n \in \omega$, then $n \subseteq \omega$, since ω is transitive. Since every subset of a well–ordered set is also well–ordered, we see that each $n \in \omega$ is also well–ordered by $\in$. Thus if either $x \in \omega$ or $x = \omega$, then x has the following properties:

(i) x is transitive.

(ii) The $\in$–relation well–orders x.

This means that every $n \in \omega$, as well as ω itself, is an *ordinal*: An ordinal is a transitive set which is well–ordered by the $\in$–relation. So we have ordinals

$$0, 1, 2, 3, \ldots, \omega$$

But now it follows easily that $S(\omega), SS(\omega), SSS(\omega), \ldots$ are transitive and well–ordered by $\in$ also. This yields *transfinite ordinals*

$$\omega + 1, \omega + 2, \omega + 3, \ldots$$

We will examine ordinals more closely in the next chapter, where we shall see that every well–ordering is order–isomorphic to a unique ordinal.

We remark that assuming the Axiom of Foundation, the $\in$–relation is a well–ordering as soon as it is a linear ordering. However, we have not used Foundation in this chapter.

3.3 Recursion

In practically all branches of mathematics there are instances of definitions by induction. Typically, one is given an initial object x_0 and then a rule to transform a given object x_n into a new object x_{n+1}. The following theorem shows that our definition of the natural numbers allows for such definitions:

Theorem 3.3.1 (Recursion Theorem — Version I)
Let $f : X \to X$ be a function, and let $x_0 \in X$. Then there exists a unique *function $g : \omega \to X$ with the properties that*

$$g(0) = x_0 \qquad g(n+1) = f(g(n)) \tag{$\star$}$$

□

The idea is that $g(0) = x_0$, $x_1 = g(1) = f(g(0)) = f(x_0)$, $x_2 = g(2) = f(g(1)) = f(x_1)$, etc.

The proof of the above theorem is given by the next exercise:

Exercise 3.3.2 (a) First prove uniqueness: If $g_1 : \omega \to X$ and , $g_2 : \omega \to X$ are two functions which both satisfy ($\star$) of Theorem 3.3.1, then $g_1 = g_2$.

(b) Let $\mathcal{F}$ be the set of all subsets of $F \subseteq \omega \times X$ with the following properties:

(i) $\langle 0, x_0 \rangle \in F$;

(ii) If $\langle n, x \rangle \in F$, then $\langle n+1, f(x) \rangle \in F$.

Let $g := \bigcap \mathcal{F}$. Explain why g is a set, and why $g \in \mathcal{F}$.

(c) Since g is a set of ordered pairs it is a binary relation. Show that $\mathrm{dom}(g) = \omega$.

(d) It remains to show that the relation g is a function, i.e. that if $\langle n, x \rangle, \langle n, y \rangle \in g$, then $x = y$. Define

$$S := \{n \in \omega : \forall x, y \in X \ (\langle n, x \rangle \in g \wedge \langle n, y \rangle \in g \to x = y)\}$$

We must show that $S = \omega$. First show that $0 \in S$.
[Hint: Suppose that $\langle 0, y \rangle \in g$ for some $y \neq x_0$. Show that $G - \{\langle 0, y \rangle\} \in \mathcal{F}$.]

(e) Finally, show that $S = \omega$.
[Hint: Show that S is inductive. If $n \in S$, there is a unique x such that $\langle n, x \rangle \in G$, and hence certainly $\langle n+1, f(x) \rangle \in G$. Suppose that $\langle n+1, y \rangle \in G$ for some $y \neq x$. Show that $G - \{\langle n+1, y \rangle\} \in \mathcal{F}$.]

□

In some definitions by induction, one needs not just the current value x_n to define x_{n+1}, but also earlier values x_m for $m < n$. For example, the Fibonacci sequence is defined by

$$x_0 = x_1 = 1 \qquad x_{n+1} = x_{n-1} + x_n \text{ for } n > 1$$

In general, we may want to define x_{n+1} as a function of $x_0, x_1, \dots, x_n$. The way to do this is as follows: Given a set X, let

$$X^{<\omega} := \{h : h \text{ is a function with } \mathrm{dom}(h) \in \omega \text{ and } \mathrm{ran}(h) \subseteq X\}$$

It is not hard to see that $X^{<\omega}$ is a set, as each $h \in X^{<\omega}$ is a subset of $\omega \times X$. If $h \in X^{<\omega}$ has $\mathrm{dom}(h) = n$, and $h(k) = h_k \in X$ for $k < n$, then we indicate this state of affairs by writing

$$h = [\![h_0, \dots, h_{n-1}]\!]$$

i.e. we may think of an element of $X^{<\omega}$ as being finite sequence of elements of X. This is merely a temporary notation to make the discussion that follows easier to understand. Note that the empty sequence $[\![\ ]\!] := \varnothing \in X^{<\omega}$.

Now let $f : X^{<\omega} \to X$ be a function. Operating on an intuitive level, we can inductively define a sequence

$$x_0 = f([\![\ ]\!]) \qquad x_1 = f([\![x_0]\!]) \qquad x_2 = f([\![x_0, x_1]\!]) \dots \qquad x_{n+1} = f([\![x_0, x_1 \dots, x_n]\!])$$

If we (still operating intuitively) define a function $g : \omega \to X : n \mapsto x_n$, then we have

$$g(0) = x_0 = f([\![\ ]\!]) = f(g \restriction 0)$$

because $g \restriction 0 = \varnothing = [\![\ ]\!]$.
Next,

$$g(1) = x_1 = f([\![x_0]\!]) = f(g \restriction 1)$$

because $g \restriction 1$ is the finite sequence $[\![g(0)]\!] = [\![x_0]\!]$.
Then,

$$g(2) = x_2 = f([\![x_0, x_1]\!]) = f(g \restriction 2)$$

because $g \restriction 2 = [\![g(0), g(1)]\!] = [\![x_0, x_1]\!]$, etc.

The following theorem states that the above construction can be accomplished within the confines of ZF$^-$:

Theorem 3.3.3 (Recursion Theorem — Version II)
For any set X and any function $f : X^{<\omega} \to X$ there is a unique *function $g : \omega \to X$ such that*

$$g(n) = f(g \restriction n) = f([\![g(0), g(1), \dots, g(n-1)]\!]) \qquad \textit{for all } n \in \omega$$

Proof: Uniqueness of g follows easily by induction. To prove that g exists, we use Theorem 3.3.1. First, define $F : X^{<\omega} \to X^{<\omega}$ by $F(h) := h \cup \{\langle \mathrm{dom}(h), f(h)\rangle\}$. This means that

$$F([\![x_0, \dots, x_{n-1}]\!]) := [\![x_0, \dots, x_{n-1}, f([\![x_0, \dots, x_{n-1}]\!])]\!]$$

By Theorem 3.3.1, there is a unique $G : \omega \to X^{<\omega}$ such that

$$G(0) = [\![\ ]\!] \qquad G(n+1) = F(G(n))$$

Thus $G(n+1) = G(n) \cup \{\langle n, f(G(n))\rangle\}$. Observe that each $G(n)$ is a function $G(n) : n \to X$, and that $G(n+1) \restriction n = G(n)$. Now let $g := \bigcup \mathrm{ran}(G) = \bigcup_{n \in \omega} G(n)$. Then $g : \omega \to X$, and $g \restriction n = G(n)$ for all $n \in \omega$. Hence $g(n) = G(n+1)(n) = f(G(n)) = f(g \restriction n)$, as required.

$\dashv$

Now that we've constructed the natural numbers, it is possible to define the operations of addition and multiplication on the set of natural numbers by recursion. However, we will leave this till a later chapter, where we shall define these operations on the class of all ordinals. We can also:

- Construct the integers $\mathbb{Z}$ as a set of equivalence classes of ordered pairs of natural numbers.

- Construct the rational numbers $\mathbb{Q}$ as a set of equivalence classes of integers.

- Construct the real numbers $\mathbb{R}$, as the set of *Dedekind cuts* (L, U), where L, U are non–empty subsets of $\mathbb{Q}$ with the property that U has no minimum element, and for all $l \in L$ and $u \in U$ we have $l < u$.

- Construct the complex numbers $\mathbb{C}$ as the set of ordered pairs of real numbers.

Of course, we must also define all the usual operations and relations on these sets as well. All this can be done, but we won't pursue it in this course. Try Edmund Landau's *Foundations of Analysis* for a mercilessly rigorous approach.

3.4 Finite and Infinite Sets

Observe that if $n \in \omega$, then $n = \{m : m < n\}$, because $<$ is $\in$. Further observe that if $m, n \in \omega$, then $m \leq n$ if and only if $m \subseteq n$, because $\in$ is transitive on the set of natural numbers.

Definition 3.4.1 Suppose that $n \in \omega$. A set X is said to have n elements if and only if there is a bijection from X to the set n. In that case we say that the *cardinality* of X is n, and write $|X| = n$. A set X is said to be *finite* if there is $n \in \omega$ such that $|X| = n$. A set is said to be *infinite* if and only if it is not finite.

□

Thus a set X has n elements if it can be indexed by the set n, i.e. can be written $X := \{x_i : i \in n\}$. Since the $\in$ is $<$, this can also be written as $X := \{x_i : i < n\}$. This state of affairs may also be indicated by writing $X := \{x_0, x_1, \dots, x_{n-1}\}$. Clearly each $n \in \omega$ is a finite set, and $|n| = n$.

Theorem 3.4.2 *(a) The notion of "number of elements of a finite set" is well-defined: If $|X| = n$ and $|X| = m$, then $n = m$.*

(b) A subset of a finite set is finite: If X is finite, and $Y \subseteq X$. Then Y is finite. In addition, $|Y| \leq |X|$.

(c) The image of a finite set is finite: If X is finite and f is a function, then $f[X]$ is finite. In addition $|f[X]| \leq |X|$.

(d) The union of a finite family of finite sets is finite.

(e) The power set of a finite set is finite.

(f) A set X is infinite if and only if for every $n \in \omega$ there is an injection $f : n \rightarrowtail X$.

(g) (Using AC) A set X is infinite if and only if there is an injection $f : \omega \rightarrowtail X$.

□

The proofs of the above assertions are contained in the exercises that follow. First, however, we state and prove a Lemma:

Lemma 3.4.3 *Let $n \in \omega$ and let X be a* proper *subset of n. Then there is no bijection $f : X \rightarrowtail\!\!\!\!\!\twoheadrightarrow n$.*

Proof: Let $m \in \omega$ be least such that there a bijection $f : m \rightarrowtail\!\!\!\!\!\twoheadrightarrow X$ onto a proper subset $X \subsetneqq m$. Then $m > 0$ (because $0 = \varnothing$ has no proper subsets), and hence by Lemma 3.2.3(d) there is $n \in \omega$ such that $m = n + 1$.

Now consider two cases:

<u>Case 1:</u> If $n \in X$, there is $k \leq n$ such that $f(k) = n$. Let $g : n + 1 \rightarrowtail\!\!\!\!\!\twoheadrightarrow n + 1$ be the bijection which has $g(k) = n$, $g(n) = k$, and which has $g(m) = m$ for all $m \notin \{k, n\}$. (Thus g is the permutation which swaps k and n, and leaves all other elements of $n + 1$ unmoved.) Then $\tilde{f} := f \circ g$ is a bijection $\tilde{f} : n + 1 \rightarrowtail\!\!\!\!\!\twoheadrightarrow X$ with $\tilde{f}(n) = n$. It is easy to see that $\tilde{f} \restriction n$ maps n

onto $X - \{n\}$, a proper subset of n. But this contradicts the minimality of $m = n + 1$.
Case 2: If $n \notin X$, then $X \subseteq n$. It is now easy to see that $f \upharpoonright n$ maps n onto $X - \{f(n)\}$, a proper subset of n — again contradicting the minimality of m.

$\dashv$

Exercise 3.4.4 (a) Prove Theorem 3.4.2(a): Show that if $|X| = n$ and $|X| = m$, then $n = m$.
[Hint: First show that if $n \neq m$, then either $n \subsetneqq m$ or else $m \subsetneqq n$. Now use Lemma 3.4.3.]

(b) Show (without using the Axiom of Choice) that the following are equivalent for $n, m \in \omega$:

(i) $n \leq m$.

(ii) There is an injection $n \rightarrowtail m$.

(iii) There is a surjection $m \twoheadrightarrow n$.

[Hint: (ii) implies (i): If $f : n \rightarrowtail m$ and $m < n$, then f is a bijection from n onto $\mathrm{ran} f \subsetneqq n$.
(iii) implies (ii): If $g : m \twoheadrightarrow n$, one can define a right inverse $f : n \to m$ without using AC.]

(c) Suppose that X, Y are finite sets. Show (without using AC) that the following are equivalent:

(i) $|X| \leq |Y|$.

(ii) There is an injection $f : X \rightarrowtail Y$.

(iii) There is a surjection $g : Y \twoheadrightarrow X$.

[Hint: Use (b).]

(d) Prove Theorem 3.4.2(b): Show that if X is finite and $Y \subseteq X$, then Y is finite, and $|Y| \leq |X|$.

(e) Prove Theorem 3.4.2(c): Suppose that X is a finite set, and that f is a function. Show that $f[X]|$ is a finite set, and that $|f[X]| \leq |X|$.

□

Exercise 3.4.5 (a) Suppose that X, Y are finite sets. Show that $X \cup Y$ is finite.
[Hint: Recall that we haven't defined addition on the natural numbers, only the successor function. Suppose that m is least for which there are finite sets X, Y with $|X| = m$ and such that $X \cup Y$ infinite. Show that $m = n + 1$ for some n. Let $x \in X$, and consider $X' := X - \{x\}$ and $Y' := Y \cup \{x\}$.]

(b) Prove Theorem 3.4.2(d): Show that if $\mathcal{X} = \{X_0, X_1, \ldots, X_{n-1}\}$ is a finite family of finite sets, then $\bigcup \mathcal{X} = \bigcup_{k=0}^{n-1} X_k$ is finite.

(c) Prove Theorem 3.4.2(e): Show that if X is finite, then $\mathcal{P}(X)$ is finite.
[Hint: Induction on $|X|$. Observe that if $x \in X$, then $\mathcal{P}(X) = \mathcal{X} \cup \mathcal{Y}$, where $\mathcal{X} := \{Y \subseteq X : x \notin Y\}$ and $\mathcal{Y} := \{Y \subseteq X : x \in Y\}$.]

□

Exercise 3.4.6 Prove Theorem 3.4.2(f): Show that a set X is infinite if and only if there is an injection $f : n \rightarrowtail X$ for every $n \in \omega$. Conclude that every inductive set is infinite.

□

A set X is said to be *Dedekind infinite* if there is a bijection from X onto a *proper subset* of X, and *Dedekind finite* otherwise.

Exercise 3.4.7 (a) Show that every finite set is Dedekind finite, and that every Dedekind infinite set is infinite.

(b) Prove (without using AC) that a set X is Dedekind infinite if and only if there is an injection $\omega \rightarrowtail X$.
[Hint: $\Rightarrow$: Suppose that $Y \subsetneqq X$ is a proper subset of X, and that $f : X \rightarrowtail Y$ is an injection. Let $x_0 \in X - Y$, and define $g : \omega \to X$ by

$$g(0) = x_0 \qquad g(n+1) = f(g(n))$$

$\Leftarrow$: If $g : \omega \rightarrowtail X$, consider $f : X \to X$ by

$$f(x) := \begin{cases} x & \text{if } x \notin \operatorname{ran}(g) \\ g(S(g^{-1}(x))) & \text{if } x \in \operatorname{ran}(g) \end{cases}$$

where S is the successor function.]

(c) Prove (with the aid of AC) that a set X is (in)finite if and only if it is Dedekind (in)finite. Conclude that Theorem 3.4.2(g) holds.
[Hint: Suppose that $\mathcal{X}$ is the set of all non–empty subsets of X, and that $F : \mathcal{X} \to \bigcup \mathcal{X}$ is a choice function. Consider $g : \omega \to X$ defined by

$$g(n) := F(X - \operatorname{ran}(g \upharpoonright n))$$

and use Theorem 3.4.2(f).]

□

Chapter 4

Ordinals

4.1 Well–Orderings

Let $(X,<)$ be a totally ordered set. If $x \in X$, then we define the *initial segment* given by x by:

$$X(x) := \{y \in X : y < x\}$$

Thus $X(x)$ is the set of all elements that are strictly below x. Clearly $X(x)$ is a totally ordered set also (with the relation $\leq\restriction X(x) \times X(x)$ inherited from X).

Lemma 4.1.1 *Suppose that $(X,\leq)$ is a total ordering. Define $\mathcal{X} := \{X(x) : x \in X\}$. Then the map $f : X \to \mathcal{X} : x \mapsto X(x)$ is an order–isomorphism from $(X,\leq)$ to $(\mathcal{X},\subseteq)$.*

Proof: It suffices to show that $x \leq y$ if and only if $X(x) \subseteq X(y)$. It is obvious from the transitivity of $<$ that $X(x) \subseteq X(y)$ whenever $x \leq y$. Conversely, suppose that $X(x) \subseteq X(y)$. If $x \not\leq y$, then $y < x$, so $y \in X(x)$, and hence $y \in X(y)$ — contradiction.

$\dashv$

When two partial orderings $(X,<)$ and $(Y,\prec)$ are order–isomorphic, we say that they have the same *order–type*.

Recall that a total ordering $(W,<)$ is called a *well–ordering* if every non–empty subset of X has a $<$–least element. Clearly, the set of natural numbers, together with its usual ordering, is a well–ordering.

The following lemma gives two methods for constructing well–orderings on products. It will prove useful when we discuss the arithmetic of ordinal and cardinal numbers.

Lemma 4.1.2 *(a) Suppose that $(X,<)$ and $(Y,\prec)$ are strict partially (totally, well–) ordered sets. Define the* lexicographic ordering[1] *$\sqsubset$ on $X \times Y$ by*

$$\langle x_1, y_1\rangle \sqsubset \langle x_2, y_2\rangle \qquad \Longleftrightarrow \qquad x_1 < x_2 \ \vee\ (x_1 = x_2 \ \ \wedge y_1 \prec y_2)$$

Then $\sqsubset$ is a strict partial(totally, well–) ordering on $X \times Y$.

[1] *Lexicography: The editing or making of a dictionary.*
In a dictionary, words are ordered lexicographically. To decide which word occurs first in the dictionary, you look at their first letters. If these are the same, you look at their second letters. . .

(b) Suppose that $(X,<)$ is a strict well–ordered set. Define the canonical ordering $\sqsubset$ *on the product $X \times X$ by*

$$\langle x,y\rangle \sqsubset \langle x',y'\rangle \qquad \Longleftrightarrow \qquad \begin{cases} & \max\{x,y\} < \max\{x',y'\} \\ \textit{or} & \max\{x,y\} = \max\{x',y'\} \textit{ and } x < x' \\ \textit{or} & \max\{x,y\} = \max\{x',y'\},\ x = x',\ \textit{and } y < y' \end{cases}$$

Then $\sqsubset$ is a strict well–ordering on $X \times X$.

□

Exercise 4.1.3 Consider $(\omega,<)$, the set of natural numbers equipped with its usual order relation.

(a) Write down the first ten elements in the lexicographic ordering $\sqsubset_l$ of $\omega \times \omega$. Hence show that $(\omega \times \omega, \sqsubset_l)$ is a well–ordering, but not order–isomorphic to $(\omega,<)$.
[Hint: Observe that the element $\langle 1,0\rangle$ is not the successor of any $\langle n,m\rangle \in \omega \times \omega$.]

(b) Prove Lemma 4.1.2(a).

(c) Write down the first ten elements in the canonical ordering $\sqsubset_c$ of $\omega \times \omega$. Hence show that $(\omega \times \omega, \sqsubset_c)$ is order–isomorphic to $(\omega,<)$, and thus a well–ordered set.

(d) Prove Lemma 4.1.2(b).

□

Exercise 4.1.4 (a) Show that a finite total ordering is a well–ordering.
[Hint: You must show that if $(X,<)$ is a finite total ordering, and if Y is a non–empty subset of X, then Y has a least element. Do this by induction on $|Y|$.]

(b) Here is a useful characterization of well–ordered sets: *A total ordering $(X,<)$ a well–ordering if and only if there is no sequence $\langle x_n\rangle$ of elements of X such that $x_1\rangle x_2 > x_3 >$* $\ldots$. Prove this (using AC).
[Hint: One direction is easy. For the other, suppose that $Y \subseteq X$ has no least element, and let F be a choice function for family the non–empty subsets of Y. Define $g : \omega \to Y$ recursively by $g(n) = F(Y - \{g(0), g(1), \ldots, g(n-1)\})$.]

□

It is precisely well–ordering that allows arguments by *induction*:

Theorem 4.1.5 (Induction on a Well–Ordering) *Let $(W,\leq)$ be a well–ordered set. Let $E \subseteq W$ be a set with the following properties.*

(i) The least element of W is a member of E.

(ii) For all $x \in X$, if $W(x) \subseteq E$, then $x \in E$.

Then $E = W$.

Proof: Else let w_0 be the least element of $W - E$. Then $W(w_0) \subseteq E$, so $w_0 \in W$—contradiction.

$\dashv$

In preparation for the definition of the transfinite ordinals, we will now begin our investigation of the structure of well–ordered sets. Most of the results that follow in this section are due to Cantor.

Theorem 4.1.6 *Let $(W, \leq)$ be a well–ordered set.*

(a) Let $f : W \to W$ be a strict order–preserving map. Then $x \leq f(x)$ for all $x \in W$.

(b) Id_W is the only order–automorphism from W to W.

Proof: (a) Let $E := \{x \in W : x \not\leq f(x)\}$. If $E \neq \varnothing$, then E has a least element x_0. Since $x_0 \not\leq f(x_0)$, we have $f(x_0) < x_0$, and hence $f(f(x_0)) < f(x_0)$. But then $f(x_0) \in E$, contradicting the fact that x_0 is the least element of E. Hence $E = \varnothing$.

(b) If f is an order–preserving automorphism, then both the maps f, f^{-1} are strict order–preserving. Hence $x \leq f(x)$ and $f(x) \leq f^{-1}(f(x))$.

$\dashv$

Corollary 4.1.7 *If $(W, \leq)$ and $(V, \preceq)$ are order–isomorphic well–ordered sets, then there is a unique order–isomorphism from W onto V.*

□

Corollary 4.1.8 *No well–ordering $(W, \leq)$ is order–isomorphic to an initial segment $W(w_0)$ of itself.*

□

Theorem 4.1.9 *Let $(W, \leq)$ and $(V, \preceq)$ be well–ordered sets. Then exactly one of the following is the case:*

(a) W, V are order–isomorphic.

(b) W is order–isomorphic to an initial segment of V, i.e. there is $v_0 \in V$ such that W is order–isomorphic to $V(v_0)$.

(c) V is order–isomorphic to an initial segment of W, i.e. there is $w_0 \in W$ such that $W(w_0)$ is order–isomorphic to V.

Proof: Define a set of ordered pairs f by

$$f := \{(w, v) \in W \times V : W(w) \text{ is order–isomorphic to } V(v)\}$$

Observe first that f is a function: For suppose that (w, v_1) and (w, v_2) both belong to f, where $v_1 \neq v_2$. Without loss of generality, we may assume that $v_1 \prec v_2$. Then $V(v_1)$ and $V(v_2)$ are order–isomorphic (as each is order–isomorphic to $W(w)$). But then, if $h : V(v_2) \to V(v_1)$ is an order–isomorphism, we must have $h(v_1) \prec v_1$, which contradicts Theorem 4.1.6.

Next observe that f is a strict order–preserving map: For suppose that $(w_1, v_1), (w_2, v_2) \in f$ are such that $w_1 < w_2$. We want to show that $v_1 \prec v_2$. Now there is an order–isomorphism $h : W(w_2) \to V(v_2)$, and this (when restricted) yields an order–isomorphism $f \restriction W(w_1) : W(w_1) \to V(h(w_1))$. Hence $(w_1, h(w_1)) \in f$, so that $h(w_1) = v_1$. But as $h(w_1) \in V(v_2)$, we see that $v_1 \prec v_2$, as required.

In particular, it follows that f is one–to–one.

Observe that $\mathrm{dom}(f)$ is *downwards–closed*: If $w_1 < w_2$ and $w_2 \in \mathrm{dom}(f)$, then $w_1 \in \mathrm{dom}(f)$ as well. Indeed, there is $v_2 \in V$ such that $W(w_2)$ is order–isomorphic to $V(v_2)$. Since $w_1 \in W(w_2)$, there must be $v_1 \in V(v_2)$ such that $W(w_1)$ is order–isomorphic to $V(v_1)$, so that $w_1 \in \mathrm{dom}(f)$.

In the same way, it can be shown that $\mathrm{ran}(f)$ is downwards–closed.

Now we distinguish three cases:

Case 1: $\mathrm{dom} f \subsetneqq W$: In that case, let w_0 be the $\leq$–least element of $W - \mathrm{dom} f$. We claim that f is an order–isomorphism from $W(w_0)$ onto V. Now because $\mathrm{dom}(f)$ is downwards–closed, there is no $w \in \mathrm{dom}(f)$ such that $w_0 < w$. Hence $\mathrm{dom}(f) = W(w_0)$. If f is not surjective, let v_0 be the $\prec$–least element of $V - \mathrm{ran}(f)$. Then clearly $W(w_0)$ is order–isomorphic to $V(v_0)$ — contradicting the fact that $w_0 \notin \mathrm{dom}(f)$. Hence f is a strictly order-preserving map with $\mathrm{dom}(f) = W(w_0)$ and $\mathrm{ran}(f) = V$.

Case 2: $\mathrm{ran} f \subsetneqq V$: In that case, let v_0 be the $\prec$–least element of $V - \mathrm{ran}(f)$. Arguing as in Case 1, it follows that f is an order–isomorphism from W onto $V(v0)$.

Case 3: $\mathrm{dom} f = W$ and $\mathrm{ran} f = V$. In that case, f is an order–isomorphism of W onto V.

$\dashv$

4.2 Ordinals

Recall that the set ω of natural numbers was shown to have the following properties:

(i) ω is a transitive set: $\forall n\ (n \in \omega \to n \subseteq \omega)$.

(ii) ω is well–ordered by the $\in$–relation: The relation $<$ on ω defined by $n < m \iff n \in m$ is a strict total–ordering on ω, and every non–empty subset of ω has a $<$-least element.

Furthermore, we saw that every $n \in \omega$ has precisely the same properties: Each $n \in \omega$ is transitive, and well–ordered by $\in$.

This suggests a definition, due to John von Neumann:

Definition 4.2.1 A set is said to be an *ordinal* if and only if it is transitive and well–ordered by $\in$.

□

Thus ω is an ordinal, as is each $n \in \omega$.

Exercise 4.2.2 Recall that $n = \{m : m < n\}$ for all $n \in \omega$. Thus $\omega(n) = n$ for all $n \in \omega$, where $\omega(n)$ is the initial segment of ω determined by n. In fact, this property *characterizes* the ordinals: Show that a strict well–ordering $(W, <)$ is an ordinal if and only if $W(x) = x$ for all $x \in W$.
[Hint: $(\Leftarrow)$: $v < w$ iff $v \in W(w)$ iff $v \in w$. If $v \in w \in W$, then $v \in W(w) \subseteq W$.]

□

We shall use the first few letters of the Greek alphabet to denote generic ordinals. Since $\in$ is a strict order relation, we shall also write $\alpha < \beta$ when $\alpha \in \beta$.

Observe that if α is a non–zero ordinal and x_0 is the *first* (i.e. the least) element of α, then $x_0 = \varnothing = 0$. For if $x \in x_0$, then $x \in \alpha$ (by transitivity of α), and $x < x_0$ (because $<$ is $\in$), which contradicts the fact that x_0 is the least element of α. Suppose now that x_1 is the *next* element in α, assuming it has one, i.e. x_1 is the least element in α which is $> x_0$. Then

$$x \in x_1 \quad\Longleftrightarrow\quad x < x_1 \quad\Longleftrightarrow\quad x = x_0$$

and hence $x_1 = \{x_0\} = \{0\} = 1$. Now if x_2 is the next element of α (again, assuming there is one), then

$$x \in x_2 \quad\Longleftrightarrow\quad x < x_2 \quad\Longleftrightarrow\quad x = x_0 \vee x = x_1$$

and so $x_2 = \{x_0, x_1\} = \{0, 1\} = 2$.

It thus appears that each ordinal α starts $0 < 1 < 2 < 3 < \dots$. Each of these elements is itself an ordinal.

In Remarks 4.2.13, we will explain that the existence of an inductive set is equivalent to the existence of an infinite set. Thus keep the Axiom of Infinity in mind while reading the proofs of the next few propositions and theorems, and observe that these proofs do not rely on this axiom.

Proposition 4.2.3 *If α is an ordinal and $\beta \in \alpha$, then β is an ordinal.*

Proof: Suppose that α is an ordinal, and that $\beta \in \alpha$. We first show that β is transitive, i.e. that if $\gamma \in \beta$, then $\gamma \subseteq \beta$. Now if $\gamma \in \beta$ and $\delta \in \gamma$, then (by transitivity of α, we have $\delta, \gamma, \beta \in \alpha$. Since $<$ is $\in$ on α, we have $\delta < \gamma < \beta$, and since $<$ is transitive, we have $\delta < \beta$, i.e. $\delta \in \beta$. Thus $\gamma \subseteq \beta$.

⊣

Now observe that every ordinal is a transitive wellfounded set: Every non–empty subset of an ordinal has a least element, which must be an $\in$–minimal element. If X, Y are transitive wellfounded sets, then a $\in$–isomorphism from X to Y is defined to be a bijective map $\pi : X \to Y$ with the property that

$$x_1 \in x_2 \qquad\Longleftrightarrow\qquad \pi(x_1) \in \pi(x_2) \qquad\qquad (x_1, x_2 \in X)$$

Note that if α, β are ordinals, then an $\in$–isomorphism from α to β is precisely an order–isomorphism.

Also observe that if $\pi : X \to Y$ is an $\in$–isomorphism between transitive wellfounded sets, and $x \in X$, then $\pi(x) = \{\pi(x') : x' \in x\} = \pi[x]$: Indeed, if $y' \in \pi(x)$, then, there is some $x' \in X$ such that $\pi(x') = y'$. It follows that $\pi(x') \in \pi(x)$, and hence that $x' \in x$.

The following proposition will be useful.

Proposition 4.2.4 *Suppose that X, Y are transitive well–founded sets, and that $\pi : X \to Y$ is an $\in$–isomorphism. Then $\pi(x) = x$ for all $x \in X$. Hence $X = Y$.*

Proof: If $\{x \in X : \pi(x) \neq x\}$ is non–empty, it has a $\in$–minimal element x_0. Now if $x \in x_0$, then also $x \in X$, and hence $\pi(x) = x$. Thus

$$\pi(x_0) = \{\pi(x) : x \in x_0\} = \{x : x \in x_0\} = x_0$$

—contradiction. Hence $\{x \in X : \pi(x) \neq x\} = \varnothing$.

$\dashv$

Theorem 4.2.5 *If α, β are ordinals, then exactly one of the following is the case:*

(i) $\alpha = \beta$, *or*

(ii) $\alpha \in \beta$, *or*

(iii) $\beta \in \alpha$.

Proof: α, β are both transitive sets well–ordered by $\in$, and hence wellfounded. Hence at most one of the above alternatives can be the case.

By Theorem 4.1.9 either (i) α, β are order–isomorphic, or (ii) α is isomorphic to an initial segment of β, or (iii) β is order–isomorphic to an initial segment of α.

Suppose that α is isomorphic to an initial segment of β, i.e. that there is $\beta_0 \in \beta$ and an order–isomorphism $\pi : \alpha \rightarrowtail\!\!\!\!\rightarrow \beta(\beta_0)$. Observe that

$$\beta(\beta_0) := \{\gamma \in \beta : \gamma < \beta_0\} = \{\gamma \in \beta : \gamma \in \beta_0\} = \beta \cap \beta_0 = \beta_0$$

because $\beta_0 \subseteq \beta$. Thus π is an $\in$–isomorphism from the transitive wellfounded set α onto the transitive wellfounded set β_0. By Proposition 4.2.4, we have $\alpha = \beta_0$, and thus $\alpha \in \beta$.

The remaining two cases can be dealt with in a similar fashion.

$\dashv$

The next corollary follows immediately:

Corollary 4.2.6 *If α, β are ordinals, then $\alpha \in \beta$ if and only if $\alpha \subsetneqq \beta$.*

□

We denote the class of all ordinals by

$$\mathbf{ON} := \{\alpha : \alpha \text{ is an ordinal}\}$$

We shall soon see that **ON** is a proper class, i.e. not a set. For the moment, however, note that **ON** is a class which is well–ordered by $\in$: By Theorem 4.2.5, $\in$ is a strict linear ordering on **ON**. To see that it is a well–ordering, suppose that $\mathbf{Y}$ is a non–empty subclass of **ON**, and let $\alpha \in \mathbf{Y}$. If $\alpha \cap \mathbf{Y} = \varnothing$, then α is the least element of $\mathbf{Y}$. Else, $\alpha \cap \mathbf{Y}$ is a non–empty subset of the ordinal α, and hence has a least element β. That β is easily seen to be the least element of $\mathbf{Y}$. Thus every non–empty subclass of **ON** has a least element.

Here are some additional facts about ordinals that are easy to establish:

Proposition 4.2.7 *(a) If α is an ordinal, then $\alpha = \{\beta : \beta < \alpha\}$.*

(b) If α, β are ordinals, then exactly one of the following is the case: either α is an initial segment of β, or $\alpha = \beta$, or β is an initial segment of α.

(c) If $\mathcal{C}$ is a non–empty class off ordinals, then $\bigcap \mathcal{C}$ is an ordinal. In fact, $\bigcap \mathcal{C}$ is the least element of $\mathcal{C}$ (i.e. $\bigcap \mathcal{C} \in \mathcal{C}$ and $\bigcap \mathcal{C} = \inf \mathcal{C}$).

(d) If $\mathcal{C}$ is a set *of ordinals, then $\bigcup \mathcal{C}$ is an ordinal. Furthermore, $\bigcup \mathcal{C} = \sup \mathcal{C}$.*

(e) If α is an ordinal, so is $S(\alpha) := \alpha \cup \{\alpha\}$. Furthermore, $S(\alpha)$ is the successor of α, i.e. the least ordinal which is $> \alpha$.

□

Exercise 4.2.8 Prove the preceding proposition.

□

Since $S(\alpha)$ is the successor of α we write $\alpha + 1 := S(\alpha)$. An ordinal α is said to be a *successor ordinal* if an only if there is β such that $\alpha = \beta + 1$. An ordinal which is not a successor ordinal is called a limit ordinal. Observe that 0 is a limit ordinal, as is ω. Indeed, ω is the smallest non–zero limit ordinal: For if $\alpha < \omega$ is a non–zero ordinal, it is a non–zero natural number, and hence a successor ordinal. Furthermore, the successor of a natural number is a natural number, so ω is not a successor ordinal.

Exercise 4.2.9 Suppose that $\alpha_0 < \alpha_1 < \alpha_2$ is a strictly increasing sequence of ordinals. Show that $\sup_{n<\omega} \alpha_n$ is a limit ordinal.

□

Exercise 4.2.10 If α is a limit ordinal, then $\alpha = \sup\{\beta : \beta < \alpha\}$. If α is a limit ordinal, then $\sup \alpha < \alpha$.

□

Exercise 4.2.11 Show that the class **ON** of all ordinals is a proper class, i.e. not a set. [Hint: If **ON** is a set, then by Proposition 4.2.7, $\sup$ **ON** is an ordinal.]

□

The result of the previous exercise — that the collection of ordinals is too big to be a set — is known as the Burali–Forti paradox, after Burali–Forti, who published it in 1897. It is probably the earliest of the paradoxes that set theory had to contend with in its initial years.

The *Axiom of Replacement* has not yet been needed in our development of set theory, but that is about to change.

Theorem 4.2.12 *Every well–ordered set is isomorphic to a unique ordinal.*

Proof: Let $(W, <)$ be an arbitrary well–ordered set. Define a binary relation F by

$$(w, \alpha) \in F \quad \Longleftrightarrow \quad w \in W \text{ and } \alpha \in \mathbf{ON} \text{ and } W(w) \text{ is order–isomorphic to } \alpha.$$

Then by Theorem 4.2.5 F is a class function. By Theorem 4.1.9, F is injective. Moreover, it is easy to see that F is strictly order–preserving.

Next, we show that if $w \in W$, then $F[W(w)]$ is an ordinal. (We are not claiming that $W(w) \subseteq \text{dom}(F)$.): Firstly, since $W(w)$ is a set, so is $F[W(w)]$, by the Axiom of Replacement. Now $F[W(w)] \neq \mathbf{ON}$, because $\mathbf{ON}$ is not a set (by Exercise 4.2.11). If γ is the least ordinal which is not an element of $F[W(w)]$, then certainly $F[W(w)] = \{\beta \in \mathbf{ON} : \beta < \gamma\} = \gamma$.

In the same way it can be shown that $F[W]$ is an ordinal.

Now we demonstrate that $\text{dom}(F) = W$: For else $W - \text{dom}(F)$ has a least element w. Then $\xi := F[W(w)]$ is an ordinal, and $F : W(w) \to \xi$ is a order–preserving bijection, and hence an order–isomorphism. Hence $(w, \xi) \in F$ — contradiction.

Thus if we define $\alpha := F[W]$, then F is an order–isomorphism from W onto α. Hence every well–ordered set is order–isomorphic to some ordinal. That ordinal must be unique, because no two distinct ordinals are order–isomorphic (by Theorem 4.1.9 and Proposition 4.2.7).

⊣

Remarks 4.2.13 Recall that the Axiom of Infinity asserted: There is an inductive set. The set ω of natural numbers is then the smallest inductive set. A set X was defined to be finite if it can be placed into bijective correspondence with some $n \in \omega$, i.e. if $|X| = n$ for some $n \in \omega$. A set was said to be infinite if it is not finite. We showed that a set X is infinite if and only if for every $n \in \omega$ there is an injection $n \rightarrowtail X$.

We now want to show that, in the theory ZFC 0–5 + Replacement, the existence of an inductive set is *equivalent* to the existence of an infinite set. First suppose that there exists an inductive set X. It follows that ω is a set (namely the intersection of the non–empty class of all inductive sets). If $n \in \omega$ then $n \subseteq \omega$, so that there is an obvious injection $n \rightarrowtail \omega$. Hence ω is infinite. This proves that if there is an inductive set, then there is an infinite set.

Conversely, suppose that there is an infinite set X. Certainly every natural number is an ordinal. But it is, at this stage, conceivable that $\mathbf{ON}$ is the class of *all* natural numbers, i.e. that $\omega = \mathbf{ON}$ is a proper class. (Obviously, if $\mathbf{ON}$ contains some ordinal $\alpha > \omega$, then $\omega \in \mathbf{ON}$ as well, by transitivity, and hence ω would be a set. So if ω is a proper class, then $\omega = \mathbf{ON}$.) Let $\mathcal{X} := \{Y \subseteq X : Y \text{ is finite}\}$. By Power Set and Separation, $\mathcal{X}$ is a set. Define a binary relation F by

$$(Y, n) \in F \qquad \Longleftrightarrow \qquad n \in \omega \wedge |Y| = n$$

Then F is clearly a class function (i.e. if $(Y, n_1) \in F$ and $(Y, n_2) \in F$, then $n_1 = n_2$). By the Axiom of Replacement, $F[\mathcal{X}]$ is a set. We claim that $F[\mathcal{X}] = \omega$. If not, there is a least $m \in \omega$ such that $m \notin F[\mathcal{X}]$, because $\omega = \mathbf{ON}$ is a well–ordered class. Now $m \neq 0$, since $\varnothing \in \mathcal{X}$. Thus $m = n + 1$ for some $n \in \omega$. Hence $n \in F[\mathcal{X}]$, i.e. there is $Y \subseteq X$ such that $|Y| = n$. But since X is infinite, we cannot have $Y = X$, and thus there is $x \in X - Y$. Then $|Y \cup \{x\}| = n + 1$, so $n + 1 \in F[\mathcal{X}]$ — contradiction. Thus $F[\mathcal{X}] = \omega$ is a set, i.e. an inductive set exists.

□

Remarks 4.2.14 Observe that since ω is an ordinal, we may inductively define new ordinals $\omega + n$ (for $n \in \omega$) by

$$\omega + 0 := \omega \qquad \omega + (n+1) = S(\omega + n)$$

The class $\omega \cup \{\omega + n : n \in \omega\}$ is clearly transitive an well–ordered by $\in$, and thus ought to be an ordinal: call it $\omega + \omega$. The problem is that the Axioms ZFC 0–7 are not strong enough to prove that $\{\omega + n : n \in \omega\}$ is a set. It is, at this stage, conceivable that $\omega + \omega = \mathbf{ON}$ is the class of all ordinals.

Consider, however, the formula $\varphi(n, y)$ which asserts that $n \in \omega$ and that $y = \omega + n$. Clearly $\varphi(n, y)$ defines a function F, in the sense that if $\varphi(n, y_1)$ and $\varphi(n, y_2)$, then $y_1 = y_2$. By the Axiom of Replacement, $F[\omega] = \{y : \exists n \in \omega\ (\varphi(n, y)\}$ is a set, i.e. $\{\omega + n : n \in \omega\}$ is a set. Then $\omega + \omega := \omega \cup \{\omega + n : n \in \omega\}$ is a set as well. Clearly the ordinal $\omega + \omega$ is a limit ordinal.

□

4.3 Transfinite Induction and Recursion

Induction is a major tool for proving results about natural numbers. Recursion is a major tool for constructing functions whose domain is the set of natural numbers. The aim of this section is to extend these tools to the class **ON**.

The Induction Principle for the natural numbers reads as follows: If $X \subseteq \omega$ is such that (i) $0 \in X$, and (ii) $n \in X \to n + 1 \in X$, then $X = \omega$.
Each non–zero natural number is a successor ordinal. In general, however, there are also limit ordinals that need to be taken care of. Hence:

Theorem 4.3.1 (Transfinite Induction) *Let* $\mathbf{X}$ *be a class of ordinals. Suppose that*

(i) $0 \in \mathbf{X}$;

(ii) If $\alpha \in \mathbf{X}$, *then it follows that* $\alpha + 1 \in \mathbf{X}$;

(iii) If α *is a limit ordinal and* $\beta \in \mathbf{X}$ *for all* $\beta < \alpha$, *then it follows that* $\alpha \in \mathbf{X}$.

Then $\mathbf{X} = \mathbf{ON}$.

Proof: If $\mathbf{X} \neq \mathbf{ON}$, there is a least ordinal α such that $\alpha \notin \mathbf{X}$. By (i), $\alpha \neq 0$. By (ii), α is not a successor ordinal. By (iii), α is not a limit ordinal — contradiction, as a limit ordinal is, by definition, an ordinal which is not a successor ordinal.

$\dashv$

Exercise 4.3.2 Let $\mathbf{X}$ be a class of ordinals. Suppose that $\forall \alpha \in \mathbf{ON}\ (\alpha \subseteq \mathbf{X}\ \to\ \alpha \in \mathbf{X})$. Show that $\mathbf{X} = \mathbf{ON}$.

□

Next, we discuss transfinite recursion. Recall that our second version of the Recursion Theorem for natural numbers read as follows: If $f : X^{<\omega} \to X$ is a function (where $X^{<\omega}$ is the set of all finite sequences in X, i.e. the set of all function from a natural number to the set X), then there is a unique function $g : \omega \to X$ such that

$$g(n) = f(g \upharpoonright n) = f(\llbracket g(0), \dots, g(n-1) \rrbracket)$$

where $\llbracket x_0, \dots, x_{n-1} \rrbracket$ is the function from the set $n := \{0, 1, \dots, n-1\}$ to the set X given by $k \mapsto x_k$. In particular, $g(0) = f(\varnothing)$, $g(1) = f(\llbracket g(0) \rrbracket)$, $g(2) = f(\llbracket g(0), g(1) \rrbracket)$, etc. Each $g(n)$ is a function of the previously defined values $g(0), \dots, g(n-1)$.

To extend this result to the transfinite realm, we introduce similar notation and terminology: If α is an ordinal, then an α–sequence (or transfinite sequence of length α) is a function with domain the set $\alpha = \{\xi : \xi < \alpha\}$. We write:

$$\llbracket x_\xi : \xi < \alpha \rrbracket \qquad \text{is the function with domain the set } \alpha = \{\xi : \xi < \alpha\} \text{ given by } \xi \mapsto x_\xi$$

Given a sequence $s := \llbracket x_\xi : \xi < \alpha \rrbracket$ and a set x, we obtain the sequence $s \cup \{\langle \alpha, x \rangle\}$ with domain $\alpha + 1$, and denote it by $\llbracket x_\xi : \xi < \alpha \rrbracket^\frown x$.

We also write $\llbracket x_\xi : \xi \in \mathbf{ON} \rrbracket$ for a class function on $\mathbf{ON}$.

Suppose now that F is a *class function*, defined on the class $\mathbf{V}^{<\mathbf{ON}}$ of all transfinite sequences[2], i.e. F assigns to each transfinite sequence of sets $\llbracket x_\xi : \xi < \alpha \rrbracket$ a set $F(\llbracket x_\xi : \xi < \alpha \rrbracket)$.

We can then create a "sequence" $\llbracket x_\xi : \xi \in \mathbf{ON} \rrbracket$ recursively as follows:

$$x_\alpha := F(\llbracket x_\xi : \xi < \alpha \rrbracket)$$

so that x_α is defined as a function of previously defined values $x_\xi, (\xi < \alpha)$. Thus:

$$x_0 := F(\varnothing), \qquad x_1 := F(\llbracket x_0 \rrbracket), \qquad x_2 := F(\llbracket x_0, x_1 \rrbracket), \qquad \dots$$

This shows how to define x_n recursively for all $n < \omega$ — and we knew how to do that already. But now we continue: By the Axiom of Replacement, $\{x_n : n < \omega\}$ is a set, and hence the function $\llbracket x_n : n < \omega \rrbracket$ exists (i.e. is a set). We thus have

$$x_\omega := F(\llbracket x_n : n < \omega \rrbracket), \qquad x_{\omega+1} := F(\llbracket x_n : n < \omega \rrbracket^\frown x_\omega), \qquad x_{\omega+2} := F(\llbracket x_n : n < \omega \rrbracket^\frown x_\omega{}^\frown x_{\omega+1}), \qquad \dots$$

In general, if $\alpha \in \mathbf{ON}$ is such that x_ξ has already been defined for all $\xi < \alpha$, then by the Axiom of Replacement, the function $\llbracket x_\xi : \xi < \alpha \rrbracket$ exists (i.e. is a set), and we may define $x_\alpha := F(\llbracket x_\xi : \xi < \alpha \rrbracket)$.

In this way we build up a class function $G : \mathbf{ON} \to \mathbf{V} : \alpha \mapsto x_\alpha$

Theorem 4.3.3 (Transfinite Recursion) *Let F be a class function (on $\mathbf{V}$). Then there is a unique function G on $\mathbf{ON}$ such that*

$$G(\alpha) = F(G \upharpoonright \alpha) \qquad \textit{for all } \alpha \in \mathbf{ON} \tag{$\star$}$$

Proof: Define a first–order formula $\psi(\alpha, h)$ which asserts that $\alpha \in \mathbf{ON}$, and h is a function $h := \llbracket x_\xi : \xi < \alpha \rrbracket$ with domain α such that

$$\text{For all } \beta < \alpha,\ h(\beta) = F(h \upharpoonright \beta), \quad \text{i.e. } x_\beta = F(\llbracket x_\xi : \xi < \beta \rrbracket)$$

[2] We include the finite sequences in the class of transfinite sequences.

We show that ψ defines a class function: If both $\psi(\alpha, h_1)$ and $\psi(\alpha, h_2)$ hold, then $h_1 = h_2$, i.e. the function h is unique (assuming it exists). Indeed, suppose that β_0 is the least ordinal $< \alpha$ such that $h_1(\beta_0) \neq h_2(\beta_0)$. Then $h_1(\beta_0) = F(\llbracket h_1(\xi) : \xi < \beta_0 \rrbracket) = F(\llbracket h_2(\xi) : \xi < \beta_0) = h_2(\beta_0)$ — contradiction.

In particular, it follows by the Axiom of Replacement that if $\varphi(\beta, h_\beta)$ holds for all $\beta < \alpha$, then $\{h_\beta : \beta < \alpha\}$ is a set.

Now let $\varphi(\alpha, x) \equiv \exists h\ [\psi(\alpha, h) \wedge x = F(h)]$. If both $\varphi(\alpha, x)$ and $\varphi(\alpha, y)$, hold, then $x = F(h) = y$, where h is the uniqe function such that $\psi(\alpha, h)$. Hence φ defines a class function

$$G(\alpha) = x \qquad \Longleftrightarrow \qquad \varphi(\alpha, x)$$

It remains to show that $\mathrm{dom}(G) = \mathbf{ON}$, i.e. that $G(\alpha)$ is defined for every $\alpha \in \mathbf{ON}$. If not, let α_0 be the least ordinal such that $\alpha_0 \notin \mathrm{dom}(G)$. Then for each $\beta < \alpha_0$, there is a unique function $h_\beta : \beta \to \mathbf{V}$ such that $\psi(\beta, h_\beta)$. We now consider two cases:

Case 1: Suppose that $\alpha_0 = \beta + 1$ is a successor ordinal. In that case there is $h_\beta = \llbracket x_\xi : \xi < \beta \rrbracket$ such that $\psi(\beta, h_\beta)$ holds. If we define $x_\beta = F(h_\beta)$, and then define $h := \llbracket x_\xi : \xi < \beta \rrbracket^\frown x_\beta$, then clearly $\psi(\alpha_0, h)$ holds also, so that $\alpha_0 \in \mathrm{dom}(G)$ after all.

Case 2: Suppose that α_0 is a limit ordinal. By the Axiom of Replacement, $\{h_\beta : \beta < \alpha_0\}$ is a set. It is easy to show that if $\eta < \beta < \alpha_0$, then $h_\eta = h_\beta \upharpoonright \alpha$. It follows that $h := \bigcup_{\beta < \alpha_0} h_\beta$ is a function with domain α_0. Furthermore, if we define

$$x_\beta := h(\beta) \qquad \text{for } \beta < \alpha_0$$

then $h = \llbracket x_\xi : \xi < \alpha_0 \rrbracket$, and $h_\beta = \llbracket x_\xi : \xi < \beta \rrbracket$ for all $\beta < \alpha_0$. Clearly, therefore, $\psi(\alpha_0, h)$ holds, and hence $\alpha_0 \in \mathrm{dom}(G)$ in this case also.

We have now shown that there exists a class function $G : \mathbf{ON} \to \mathbf{V}$ with the required property $(\star)$. Suppose that G' is another. If $G \neq G'$, let α be the smallest ordinal such that $G(\alpha) \neq G'(\alpha)$. Then $G \upharpoonright \alpha = G' \upharpoonright \alpha$, and hence $G(\alpha) = F(G \upharpoonright \alpha) = F(G' \upharpoonright \alpha) = G'(\alpha)$ — contradiction. Hence $G = G'$, i.e. there is a uniqe function with the property $(\star)$.

⊣

4.4 Ordinal Arithmetic

4.4.1 Limits

Definition 4.4.1 (a) A transfinite sequence $\llbracket \gamma_\xi : \xi < \alpha \rrbracket$ of ordinals is said to be *strictly increasing* if

$$\eta < \xi \qquad \text{implies} \qquad \gamma_\eta < \gamma_\xi$$

(b) If $\lambda > 0$ is a limit ordinal, then the *limit* of a strictly increasing sequence $\llbracket \gamma_\xi : \xi < \lambda \rrbracket$ of ordinals is defined by

$$\lim_{\xi \to \lambda} \gamma_\xi := \sup\{\gamma_\xi : \xi < \lambda\} = \bigcup_{\xi < \lambda} \gamma_\xi$$

(This limit exists, by Proposition 4.2.7.)

(c) A strictly increasing sequence $\llbracket \gamma_\xi : \xi < \alpha \rrbracket$ (or $\llbracket \gamma_\xi : \xi \in \mathbf{ON} \rrbracket$) of ordinals is said to be *normal* if it is "continuous", i.e. if for for every limit ordinal λ in its domain we have

$$\gamma_\lambda = \lim_{\xi \to \lambda} \gamma_\xi \qquad \text{i.e.} \quad \lim \gamma_\xi = \gamma_{\lim \xi}$$

□

Similarly, we say that a class function $F : \mathbf{ON} \to \mathbf{ON}$ is normal if it is strictly increasing and continuous. Of course, any sequence is a function.

Proposition 4.4.2 *Suppose that* $[\![\gamma_\xi : \xi \in \mathbf{ON}]\!]$ *is a normal sequence of ordinals.*

(a) $\gamma_\xi \geq \xi$ *for all* $\xi \in \mathbf{ON}$.

(b) γ_λ *is a limit ordinal if* λ *is a limit ordinal.*

(c) The sequence has arbitrarily large fixed points*: For every* $\alpha \in \mathbf{ON}$ *there is* $\beta > \alpha$ *such that* $\gamma_\beta = \beta$.

□

Exercise 4.4.3 Prove Proposition 4.4.2.
[Hint for (c): Define $\beta_0 := \gamma_\alpha, \beta_{n+1} := \gamma_{\beta_n}, \beta := \lim_{n\to\omega} \beta_n$.]

□

4.4.2 Addition

Definition 4.4.4 (Ordinal Addition) *We define a class function* $+ : \mathbf{ON} \times \mathbf{ON} \to \mathbf{ON}$ *by transfinite recursion as follows. Let* $\alpha \in \mathbf{ON}$.

(i) $\alpha + 0 := \alpha$.

(ii) $\alpha + (\beta + 1) := (\alpha + \beta) + 1$ *for successor ordinals* $\beta + 1$.

(iii) $\alpha + \beta := \lim_{\xi \to \beta} (\alpha + \xi)$ *for limit ordinals* $\beta > 0$.

□

Observe that, for each $\alpha \in \mathbf{ON}$, we have defined above a function $S_\alpha : \mathbf{ON} \to \mathbf{ON} : \beta \mapsto \alpha + \beta$. It is easy to see that each $S_\alpha(\cdot)$ is a normal function (i.e. strictly increasing and continuous). It is also easy to see that ordinal addition coincides with the usual addition on the set of natural numbers.

Furthermore, ordinal addition is associative:

Lemma 4.4.5 *For all* $\alpha, \beta, \gamma \in \mathbf{ON}$*, we have*

$$\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma$$

Proof: By induction on γ: For all ordinals α, β we have

$$\alpha + (\beta + 0) = \alpha + \beta = (\alpha + \beta) + 0$$

Next, if γ is such that $\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma$ holds for all α, β, then

$$\alpha + (\beta + (\gamma + 1)) = \alpha + ((\beta + \gamma) + 1) = (\alpha + (\beta + \gamma)) + 1 = ((\alpha + \beta) + \gamma) + 1 = (\alpha + \beta) + (\gamma + 1)$$

and hence the result holds for successor ordinals $\gamma + 1$.

Finally if γ is limit, and $\alpha + (\beta + \eta) = (\alpha + \beta) + \eta$ for all $\eta < \gamma$, then

$$\alpha + (\beta + \gamma) = \alpha + \lim_{\eta < \gamma} (\beta + \eta) = \lim_{\eta < \gamma} (\alpha + (\beta + \eta)) = \lim_{\eta < \gamma} ((\alpha + \beta) + \eta) = (\alpha + \beta) + \gamma$$

and hence the result holds for limit ordinals.

$\dashv$

However, ordinal addition fails to be commutative! For example:

$$1+\omega = \lim_{n<\omega}(1+n) = \omega \qquad \omega+1 := S(\omega) > \omega$$

so $1+\omega \neq \omega+1$.

Consider now $\omega+1$. This is the least ordinal which is $> \omega$. Hence it is order–isomorphic to the set of natural numbers, followed by a new maximum element.
Similarly, $\omega+2 = (\omega+1)+1$ has another additional element added to the top: It is order–isomorphic to ω followed by two new elements.
Next, consider $\omega+\omega = \lim_{n<\omega}(\omega+n)$. This is the least ordinal which is $> \omega+n$ for every $n<\omega$. Clearly, therefore, $\omega+\omega$ is order–isomorphic to ω, followed by another copy of ω.

We make this precise.

Suppose that $(X, \prec_X)$ and $(Y, \prec_Y)$ are *disjoint* totally ordered sets. We can form their *linear sum* $(X \cup Y, \prec)$ by gluing X, Y together in such a way that every element of X precedes every element of Y. This means that the linear sum of X, Y is order–isomorphic to an ordering consisting of X followed by Y.

Thus we define $\prec$ on $X \cup Y$ as follows: If $u, v \in X \cup Y$ then

$$u \prec v \qquad \Longleftrightarrow \qquad \begin{cases} \quad u, v \in X \text{ and } u \prec_X v \\ \text{or } u, v \in Y \text{ and } u \prec_Y v \\ \text{or } u \in X \text{ and } v \in Y \end{cases}$$

Observe that $\prec \, = \, \prec_X \cup \prec_Y \cup (X \times Y)$.

If $(X, \prec_X), (Y, \prec_Y)$ are totally ordered sets, but $X \cap Y \neq \varnothing$, then we can make them disjoint by a standard trick. Let $\bar{X} := \{0\} \times X$, and define $\prec_{\bar{X}}$ by $\langle 0, x\rangle \prec_{\bar{X}} \langle 0, x'\rangle$ if and only if $x \prec_X x'$. Clearly $(X, \prec_X)$ and $(\bar{X}, \prec_{\bar{X}})$ are order–isomorphic.
Similarly, let $\bar{Y} := \{1\} \times Y$ and define $\prec_{\bar{Y}}$ by $\langle 1, y\rangle \prec_{\bar{Y}} \langle 1, y'\rangle$ if and only if $y \prec_Y y'$. Then $\bar{X}, \bar{Y}$ are disjoint and order isomorphic to X, Y.
Observe that the linear sum of $\bar{X}, \bar{Y}$ is a sort of *lexicographic* ordering: For $i, j \in \{0, 1\}$ and $u, v \in X \cup Y$ we have:

$$\langle i, u\rangle < \langle j, v\rangle \qquad \Longleftrightarrow \qquad i < j \;\vee\; (i = j \wedge u < v)$$

where you have to keep track to which set each $<$ belongs.

Proposition 4.4.6 *If $\alpha, \beta \in$* **ON***, then $\alpha+\beta$ is order–isomorphic to the linear sum of α and β (in that order).*

Proof: By induction on β.

$\dashv$

Exercise 4.4.7 Let $\alpha, \beta, \gamma \in$ **ON**.

(a) Show that if $\beta < \gamma$, then $\alpha+\beta < \alpha+\gamma$.
Also find an example where $\beta < \gamma$, but $\beta+\alpha = \gamma+\alpha$.

(b) Show that if $\alpha < \beta$, then there is a unique ordinal γ so that $\alpha + \gamma = \beta$.
Also find an example where $\alpha < \beta$, but the equation $x + \alpha = \beta$ has many solutions x.
[Hint: The set $\{\xi : \alpha \leq \xi < \beta\}$ is well–ordered, so is order–isomorphic to some ordinal γ.]

□

4.4.3 Multiplication

Definition 4.4.8 (Ordinal Multiplication) *We define a class function* $\cdot : \mathbf{ON} \times \mathbf{ON} \to \mathbf{ON}$ *by transfinite recursion as follows. Let* $\alpha \in \mathbf{ON}$*.*

(i) $\alpha \cdot 0 := 0$.

(ii) $\alpha \cdot (\beta + 1) := \alpha \cdot \beta + \alpha$ *for successor ordinals* $\beta + 1$.

(iii) $\alpha \cdot \beta := \lim_{\xi \to \beta} (\alpha \cdot \xi)$ *for limit ordinals* $\beta > 0$.

□

Observe that, for each $\alpha \in \mathbf{ON}$, we have defined above a function $P_\alpha : \mathbf{ON} \to \mathbf{ON} : \beta \mapsto \alpha \cdot \beta$. It is easy to see that each $P_\alpha(\cdot)$ is a normal function (i.e. strictly increasing and continuous). It is also easy to see that $\cdot$ coincides with the usual multiplication on the set of natural numbers.

Furthermore, ordinal multiplication is associative:

Lemma 4.4.9 *For all* $\alpha, \beta, \gamma \in \mathbf{ON}$*, we have*

$$\alpha \cdot (\beta \cdot \gamma) = (\alpha \cdot \beta) \cdot \gamma$$

□

Exercise 4.4.10 Prove the preceding Lemma.

□

Observe that

$$2 \cdot \omega = \lim_{n < \omega} 2 \cdot n = \omega \qquad \omega \cdot 2 = \omega \cdot (1 + 1) = \omega \cdot 1 + \omega = \omega + \omega$$

Thus ordinal multiplication also fails to be commutative: $2 \cdot \omega \neq \omega \cdot 2$.

Indeed $2 \cdot \omega$ looks like ω–many copies of the total ordering 2: Each element of ω is replaced by a copy of 2.
On the other hand, $\omega \cdot 2$ looks like 2 copies of ω: Each element of 2 is replaced by a copy of ω.
We make this precise:

Suppose that $(X, \prec_X)$ and $(Y, \prec_Y)$ are totally ordered sets. The *linear product* of these sets is the set $X \times Y$ equipped with the *reverse lexicographical ordering* $\prec$, i.e.

$$\langle x, y \rangle \prec \langle x', y' \rangle \qquad \Longleftrightarrow \qquad \begin{cases} & y \prec_Y y' \\ \text{or} & y = y' \text{ and } x \prec_X x' \end{cases}$$

Thus, if we keep $y \in Y$ fixed, the subset $X_y := \{\langle x, y\rangle : x \in X\} \subseteq X \times Y$ is a copy of X. We see furthermore that if $y \prec_Y y'$, then each element of the copy X_y precedes each element of the copy $X_{y'}$ in the reverse lexicographic ordering. Thus, roughly, $(X \times Y, \prec)$ consists of "Y copies of X".

Proposition 4.4.11 *If $\alpha, \beta \in \mathbf{ON}$, then $\alpha \cdot \beta$ is order–isomorphic to the linear product of α and β (in that order).*

Proof: By induction on β.

$\dashv$

Exercise 4.4.12 Let $\alpha, \beta, \gamma \in \mathbf{ON}$.

(a) Show that $\alpha \cdot (\beta + \gamma) = \alpha \cdot \beta + \alpha \cdot \gamma$.
Also find an example where $(\alpha + \beta) \cdot \gamma \neq \alpha \cdot \gamma + \beta \cdot \gamma$.

(b) Show that if $\beta < \gamma$ and $\alpha > 0$, then $\alpha \cdot \beta < \alpha \cdot \gamma$.
Also find an example where $\beta < \gamma$ and $\alpha > 0$, but $\beta \cdot \alpha = \gamma \cdot \alpha$.

□

4.4.4 Exponentiation

Definition 4.4.13 (Ordinal Exponentiation) *We define a class function $\mathbf{ON} \times \mathbf{ON} \to \mathbf{ON}$ by transfinite recursion as follows. Let $\alpha \in \mathbf{ON}$.*

(i) $\alpha^0 := 1$.

(ii) $\alpha^{\beta+1} := \alpha^\beta \cdot \alpha$ *for successor ordinals* $\beta + 1$.

(iii) $\alpha^\beta = \lim_{\xi \to \beta} \alpha^\xi$ *for limit ordinals* $\beta > 0$.

□

Observe that, for each $\alpha \in \mathbf{ON}$, we have defined above a function $E_\alpha : \mathbf{ON} \to \mathbf{ON} : \beta \mapsto \alpha^\beta$. It is easy to see that each $E_\alpha(\cdot)$ is a normal function (i.e. strictly increasing and continuous). It is also easy to see that this function coincides with the usual exponentiation on the set of natural numbers.

Exercise 4.4.14 Show that if $\alpha, \beta, \gamma \in \mathbf{ON}$, then

$$\alpha^{\beta+\gamma} = \alpha^\beta \cdot \alpha^\gamma \qquad (\alpha^\beta)^\gamma = \alpha^{\beta\cdot\gamma}$$

□

4.5 Project: Goodstein's Theorem

Goodstein's Theorem is a a result about the natural numbers involving just the basic arithmetic operations. To state it, we introduce some defintions and terminology.

We describe here the procedure for writing a natural number in "superbase" form. Given a natural number $n \geq 2$, every $m \in \omega$ has a unique representation in base n, i.e. there are unique $k_0, \ldots, k_d \in \{0, \ldots, n-1\}$ such that

$$m = \sum_{i=0}^{d} k_i \cdot n^i$$

Each exponent i in the n^i–term of the above representation may be an arbitrary natural number. In particular, it is possible that $i > n$. In that case, we can write i itself in base n, i.e. $i = \sum_{j=1}^{s} l_j \cdot n^j$. If some of the exponents j are $> n$, one repeats the procedure. Clearly, after finitely many steps, the number m will have an expression in which no numbers $> n$ appear. This is the representation of m in superbase n.

For example, if we we want to write the number $m = 1\,026$ in superbase form, with base $n = 2$. Then

$$1\,026 = 2^{10} + 2^1 = 2^{2^3+2^1} + 2^1 = 2^{2^{2^1+1}+2^1} + 2^1$$

No numbers > 2 appear in this representation of $1\,026$ in superbase 2.

Problem 1. Write 59 106 in superbase 3.

□

Now consider the *Goodstein algorithm* for generating a sequence of natural numbers $(m_n)_n$ starting from an arbitrary natural number m:

Step 1. Let $m_1 := m$.

Step 2. Write m_1 in superbase 2.
Replace all occurrences of the symbol '2' by the symbol '3', to get a new number.
Subtract 1 from this new number, to get the number m_2.

Step 3. Write m_2 in superbase 3.
Replace all occurrences of the symbol '3' by the symbol '4', to get a new number.
Subtract 1 from this new number, to get the number m_3.

Step 4. Write m_3 in superbase 4.
Replace all occurrences of the symbol '4' by the symbol '5', to get a new number.
Subtract 1 from this new number, to get the number m_4.

Step 5. $\vdots$

We illustrate this for $m = m_1 = 5$.

Step 1. $m_1 = 2^2 + 1$, so $m_2 = 3^3 + 1 - 1 = 27$.

Step 2. $m_2 = 3^3$, so $m_3 = 4^4 - 1 = 255$.

Step 3. $m_3 = 3 \cdot 4^3 + 3 \cdot 4^2 + 3 \cdot 4 + 3$. Hence $m_4 = 3 \cdot 5^3 + 3 \cdot 5^2 + 3 \cdot 5 + 3 = 467$.

Continuing, we get 775, 1 197, 1 751, 2 454, 3 325, 4 382, 5 643, 7 126, 8 849, 10 830, 13 087, 15 637, 18 499, 21 691, 25 231, ...

Theorem (Goodstein's Theorem)
The sequence $(m_n)_n$ will eventually reach the number 0, for any starting value $m = m_1$.

□

The above theorem was proved in 1944. The remarkable fact about Goodstein's Theorem is this: *Even though it is a statement involving only the basic arithmetic of the natural numbers, it requires transfinite ordinals in order to prove it.* In 1981, Baker and Paris proved that Goodstein's theorem implies the consistency of first–order Peano Arithmetic. By Gödel's Second Incompleteness Theorem, PA is not strong enough to prove its own consistency. Hence Goodstein's Theorem is not provable in PA. It follows that transfinite methods are necessary even for proving results in number theory! Goodstein's theorem is unprovable without recourse to the transfinite.

We now formalize the Goodstein algorithm.

Definition: Given a base $n \geq 2$, define the function $S_n : \omega \to \omega$ recursively by

$$S_n(0) := 0 \qquad S_n\Big(\sum_{i=0}^{d} k_i \cdot n^i\Big) := \sum_{i=0}^{d} k_i \cdot (n+1)^{S_n(i)}$$

Then, for each $n \geq 1$, define functions $g_n : \omega \to \omega$ recursively by

$$g_1(m) := m \qquad g_{n+1}(m) := S_{n+1}(g_n(m)) - 1$$

□

Problem 2: Show by induction that the function S_n effects the replacing of n by $n+1$, so that

$$S_n(k \cdot n^i) = \begin{cases} k \cdot (n+1)^i & \text{if } i < n \\ k \cdot (n+1)^{n+1} & \text{if } i = n \end{cases}$$

etc.

Conclude that the sequence $(g_n(m))_n$ is precisely the Goodstein sequence starting with m.

□

Problem 3.

(a) Calculate the first few terms $m_1, m_2, \dots, m_{100}, \dots$ of the Goodstein sequence $(g_n(5))_n$ starting with $m_1 = 5$. (Play around, and push it as far as the recursive depth permitted by your programming language will allow. Observe that the terms don't seem to be converging to 0....)

(b) We now try to understand how and why the sequence m_n will eventually reach 0.

  (i) Suppose that, for some m, it is the case that $g_n(m) = n = 0 \cdot (n+1)^1 + n \cdot (n+1)^0$ in base $n+1$. For which k is $g_k(m) = 0$?

  (ii) Suppose that, for some m, it is the case that $g_n(m) = 2 \cdot (n+1)^1 + 0 \cdot (n+1)^0$. For which k is $g_k(m) = 0$?

  (iii) Suppose that, for some m, it is the case that $g_n(m) = n \cdot (n+1)^1 + n \cdot (n+1)^0$. For which k is $g_k(m) = 0$?

(c) Calculate how long it will take for $(m_n)_n$ to reach zero when $m = 4$.
Hint: There's doubling at work: $m_2 = 2 \cdot 3^2 + 2 \cdot 3 + 2$. Then $m_5 = 2 \cdot 6^2 + 1 \cdot 6 + 5$ has linear coefficient reduced by 1. Then $m_{11} = 2 \cdot 12^2 + 0 \cdot 12 + 11$ has linear coefficient reduced once more. Then $m_{23} = 1 \cdot 24^2 + 23 \cdot 24 + 23$ has the quadratic coefficient reduced by 1. Then m_{47} will have the linear coefficient reduced from 23 to 22. Doubling 22 more times, the quadratic coefficient of m_n will disappear when $n = 1 + 3 + 6 + 12 + 24 \cdots + 3 \cdot 2^{26}$. In that case $m_{n+1} = (n+1) \cdot (n+1)^1 + (n+1) \cdot (n+1)^0$. Double $(n+1)$ more times to remove the linear coeficient altogether...]

□

To prove that n_k eventually reaches zero, we use transfinite ordinal arithmetic. For $n \geq 2$, define a function $\Psi_n : \omega \to \mathbf{ON}$ as follows:

$$\Psi_n(0) := 0 \qquad\qquad \Psi_n\Big(\sum_{i=0}^{d} k_i \cdot n^i\Big) := \sum_{i=0}^{d} k_i \cdot \omega^{\Psi_n(i)}$$

For example, $\Psi_2(10) = \Psi_2(2^3 + 2^1) = \omega^{\Psi_2(3)} + \omega^{\Psi_2(1)} = \omega^{\Psi(2^1+2^0)} + \omega^{\Psi_2(2^0)} = \omega^{\omega^{\Psi_2(1)} + \omega^{\Psi_2(0)}} + \omega^{\Psi_2(0)} = \omega^{\omega+1} + 1$

Problem 4:

(a) Show that for all $2 \leq n, m < \omega$,

$$\Psi_n(m) = \Psi_{n+1}(S_n(m))$$

(b) Show that each Ψ_n is a strictly increasing function, i.e. that $\Psi_n(m+1) > \Psi_n(m)$ for all $2 \leq n, m < \omega$.

(c) Show that if $2 \leq n, m < \omega$ and $g_n(m) > 0$, then

$$\Psi_{n+2}(g_{n+1}(m)) < \Psi_{n+1}(g_n(m))$$

(d) Now prove Goodstein's Theorem: $\forall m \, \exists n \, (g_n(m) = 0)$.

□

Chapter 5

Cardinals

5.1 Cardinality: Basic Definitions

Ordinals are for *counting* (i.e. *enumerating*) sets. Such an enumeration automatically induces a well–ordering on a set X: Here is the first element x_0, here is the next x_1,..., here is the the next x_ω,.... We do not yet know whether or not every set X *can* be enumerated in such a fashion. If X is well–orderable (i.e. if there exists a well–ordering relation on X), then X can be thus enumerated, because every well–ordering is order–isomorphic to a unique ordinal. However, even when we do know that a set X can be well–ordered, the well–order induced by the enumeration of its elements is not generally unique (unless X is finite). For example, the set ω of natural numbers can be enumerated in the usual fashion:

$$0, 1, 2, 3, \ldots \tag{1}$$

It can also be enumerated by first listing all the even numbers, and then all the odd numbers:

$$0, 2, 4, \ldots, 1, 3, 5, \ldots \tag{2}$$

We can enumerate it also by first listing all the positive numbers which are not powers of some prime, followed by all the prime powers, followed by zero, as follows:

$$1, 6, 10, 12, 15, \ldots, 2, 4, 8, 16, \ldots, 3, 9, 27, 81, \ldots, 5, 25, 125, \ldots, 7, 49, \ldots \ \ldots 0 \tag{3}$$

The order–type of the above enumerations are: (1) ω, (2) $\omega + \omega$, and (3) $\omega \cdot \omega + 1$.

Obviously, the size of the set of natural numbers does not depend on how we enumerate it. To measure the size of a set, we therefore need a different notion, namely that of *cardinality.* The idea is that two sets A, B have the same size if there is a correspondence between the elements of A and B, such that to each element $a \in A$ there corresponds a unique element $b \in B$, and vice versa.

Definition 5.1.1 (a) We say that two sets A, B have the same *cardinality* (or the same *cardinal number*) and write

$$|A| = |B| \quad \Longleftrightarrow \quad \text{there exists a bijection } A \rightarrowtail\!\!\!\!\!\rightarrow B$$

In that case we say that A, B are *equivalent*, or *equipollent*, or have the *same power.*

(b) We say that the cardinality of a set A is smaller than the cardinality of B, and write

$$|A| \leq |B| \quad \iff \quad \text{there exists an injection } A \rightarrowtail B$$

We also write $|A < |B|$ to mean $|A| \leq |B|$ but $|A| \neq |B|$.

□

Recall that in the chapter in natural numbers, we wrote $|X| = n$ if there is a bijection from X to the set n. In particular $|n| = n$ for all $n \in \omega$. We also showed that $|n| = |m|$ if and only if $n = m$. Thus we may define the *finite cardinal numbers* to be the natural numbers. For infinite sets we do not yet have a candidate for the object $|X|$, i.e. the symbol $|X|$ is (at this stage) meaningless when taken in isolation. We have given meaning to statements of the form $|X| = |Y|$ and $|X| \leq |Y|$, where $|X|$ does not occur in isolation. Thus a statement such as $|X| = |X|$, though quite easy to prove, does not follow trivially from the logical rules that govern the symbol $=$. What we are asserting when we say that $|X| = |X|$ is not that the object $|X|$ is equal to itself — we do not yet have an object $|X|$ — but that there is a bijection from X to X. Equality doesn't come into it at all.

Cardinal numbers $|X|$ as definite objects *can* be defined. However, such a definition requires either the Axiom of Choice (in which case $|X|$ is the least ordinal which is able to enumerate X), or else the Axiom of Foundation (in which case $|X|$ is the family of all sets of minimal rank which are equivalent to X — the definition of *rank* requires the Axiom of Foundation). In this chapter we avoid the use of both these axioms.

The next exercise shows that the relations defined above behave as their notation suggests.

Exercise 5.1.2 (a) Show that equivalence of sets is an equivalence relation on the class of all sets:

(i) $|A| = |A|$;

(ii) If $A| = |B|$, then $|B| = |A|$;

(iii) If $|A| = |B|$, and $|B| = |C|$, then $|A| = |C|$.

(b) Show that $|A| \leq |A|$.

(c) Show that if $|A| = |A'|$ and $|B| = |B'|$, then $|A| \leq |B|$ if and only if $|A'| \leq |B'|$.

(d) Show that if $|A| \leq |B|$ and $|B| \leq |C|$, then $|A| \leq |C|$.

□

We say that a set X is *countable* if and only if $|X| \leq |\omega|$, i.e. if and only if there is an injection $X \rightarrowtail \omega$. Clearly, each natural number is countable, as is ω itself. A set which is not countable is said to be *uncountable*.

Exercise 5.1.3 (a) Show (without using (AC)) that a set X is countable if and only if there is a surjection $\omega \twoheadrightarrow X$.

Thus X is countable if and only if it can be enumerated by natural numbers, possibly with repetitions

$$X := \{x_n : n \in \omega\}$$

(b) Show that a subset of a countable set is countable.

(c) Show that X is countable if and only if either X is finite (i.e. $|X| = n$ for some $n \in \omega$) or $|X| = |\omega|$. Thus we cannot have

$$n < |X| < |\omega| \qquad \text{for all } n \in \omega$$

[Hint: Let $\omega \overset{g}{\twoheadrightarrow} X$, where X is infinite. Define $h : \omega \to X$ recursively by

$$h(n) = g(\min\{k \in \omega : g(k) \notin \operatorname{ran}(h \restriction n)\})$$

and show that h is a bijection.]

(d) Show that the union of two countable sets is countable. Conclude that the union of a finite family of countable sets is countable.[1]

(e) Show that the cartesian product of two countable sets is countable. Conclude that the cartesian product of a finite family of countable sets is countable.
[Hint: First show that $\omega \times \omega \to \omega : (i, j) \mapsto 2^i \cdot (2j + 1) - 1$ is a bijection. (You may assume that the usual laws of arithmetic hold.)]

□

Intuition for the sizes of sets may also suggest to you that $\leq$ is antisymmetric, i.e. that if $|A| \leq |B|$ and $|B| \leq |A|$, then $|A| = |B|$. However, this is not at all obvious: One has to show that if there is an injection from A to B and an injection from B to A, then there is a bijection from A to B. Nevertheless, we will proceed to show that it is the case. The proof of this fact is the first (in these notes) that deserves the appellation "beautiful".

Theorem 5.1.4 (Schröder–Bernstein[2] Theorem)
If A, B are sets such that $|A| \leq |B|$ and $|B| \leq |A|$, then $|A| = |B|$.

Proof: We first reduce the problem to the situation where $B \subseteq A$. Suppose that $A \overset{i}{\rightarrowtail} B$ and $B \overset{j}{\rightarrowtail} A$ are injections. Define $\bar{B} := j[B]$, and $A_1 := (j \circ i)[A]$. Then

$$A_1 \subseteq \bar{B} \subseteq A, \qquad |A_1| = |A|, \quad |\bar{B}| = |B|$$

If we can show that $|A| = |\bar{B}|$, then also $|A| = |B|$.

We have thus reduced the problem to the following situation: Given that $A_1 \subseteq B \subseteq A$ and that $|A_1| = |A|$, we have to show that $|A| = |B|$, i.e. that there exists a bijection $A \overset{f}{\rightarrowtail\!\!\!\!\twoheadrightarrow} B$.

Now since $|A| = |A_1|$, there is a bijection $A \overset{g}{\rightarrowtail\!\!\!\!\twoheadrightarrow} A_1$, so that $A_1 = g[A]$. Now define two sequences of sets $[\![A_n : N < \omega]\!]$ and $[\![B_n : n < \omega]\!]$ by induction as follows:

$$A_0 := A \qquad A_{n+1} := g[A_n] \qquad\qquad B_0 := B \quad B_{n+1} := g[B_n]$$

[1] You may know that the union of a countable family of countable sets is countable, but to prove that, you need (AC).

[2] Also known as the Cantor–Bernstein Theorem.

Since g is a bijection one can think of $A_0, A_1, A_2, \dots$ as copies of the same set A. Similarly $B_0, B_1, B_2, \dots$ are copies of B. A peek at Figure 5.1 may clarify matters.

Moreover,

$$A_0 \supseteq B_0 \supseteq A_1 \supseteq B_1 \supseteq \cdots \supseteq A_n \supseteq B_n \supseteq A_{n+1} \supseteq B_{n+1} \supseteq \dots$$

When we form set differences of successive sets, we obtain disjoint sets

$$A_0 - B_0, \quad B_0 - A_1, \quad A_1 - B_1, \quad B_1 - A_2, \quad \dots \quad A_n - B_n, \quad B_n - A_{n+1}, \quad \dots$$

Thus we can write A, B as disjoint unions of these set differences.

$$A = \bigcup_{0 \le n < \omega} (A_n - B_n) \cup \bigcup_{0 \le n < \omega} (B_n - A_{n+1}) \qquad B = \bigcup_{1 \le n < \omega} (A_n - B_n) \cup \bigcup_{0 \le n < \omega} (B_n - A_{n+1})$$

Now observe that g (suitably restricted) is a bijection from each set difference to the next one of the same type. To be precise,

$$g[A_n - B_n] = A_{n+1} - B_{n+1} \qquad g[B_n - A_{n+1}] = B_{n+1} - A_{n+2}$$

We can now construct the required bijection $A \overset{f}{\rightarrowtail\!\!\!\twoheadrightarrow} B$. The basic idea is to use g to move each point $x \in A_n - B_n$ to the corresponding point $f(x) \in A_{n+1} - B_{n+1}$. Points $x \in B_n - A_{n+1}$ are left unmoved, however. Thus we define

$$f(x) = \begin{cases} g(x) & \text{if } x \in \bigcup_{0 \le n < \omega} (A_n - B_n) \\ x & \text{if } x \in \bigcup_{0 \le n < \omega} (B_n - A_{n+1}) \end{cases}$$

Now observe that f is a bijection from $\bigcup_{0 \le n < \omega}(A_n - B_n)$ onto $\bigcup_{1 \le n < \omega}(A_n - B_n)$. It clearly is also a bijection from $\bigcup_{0 \le n < \omega}(B_n - A_{n+1})$ onto itself. It follows easily that $f : A \rightarrowtail\!\!\!\twoheadrightarrow B$ is a bijection.

⊣

Remarks 5.1.5 With the Schröder–Bernstein Theorem in place, we see that $\le$ defines a partial ordering on the equivalence classes of $=$. Another intuitively plausible fact is the *Law of Trichotomy*, which asserts that $\le$ induces a total ordering: If A, B are sets, then

$$\text{Either} \quad |A| < |B| \quad \text{or} \quad |A| = |B| \quad \text{or} \quad |A| > |B|$$

Cantor assumed that the Law of Trichotomy held, but was unable to supply a proof. In fact, the Law of Trichotomy is equivalent to (AC) as we shall see in a later chapter.

Recall that a set X is Dedekind infinite if and only if there is a bijection from X onto a proper subset of X. We have seen in an exercise that every Dedekind infinite set is infinite. Furthermore, we saw that a set X is Dedekind infinite if and only if there is an injection $\omega \rightarrowtail X$, i.e. if and only if $|\omega| \le |X|$. To prove that an infinite set is Dedekind infinite required (AC) however. Thus without (AC) it is possible that $|X| > n$ for all $n \in \omega$, yet $|X| \not\ge |\omega|$. In that case X, ω are incomparable elements in the ordering $\le$.

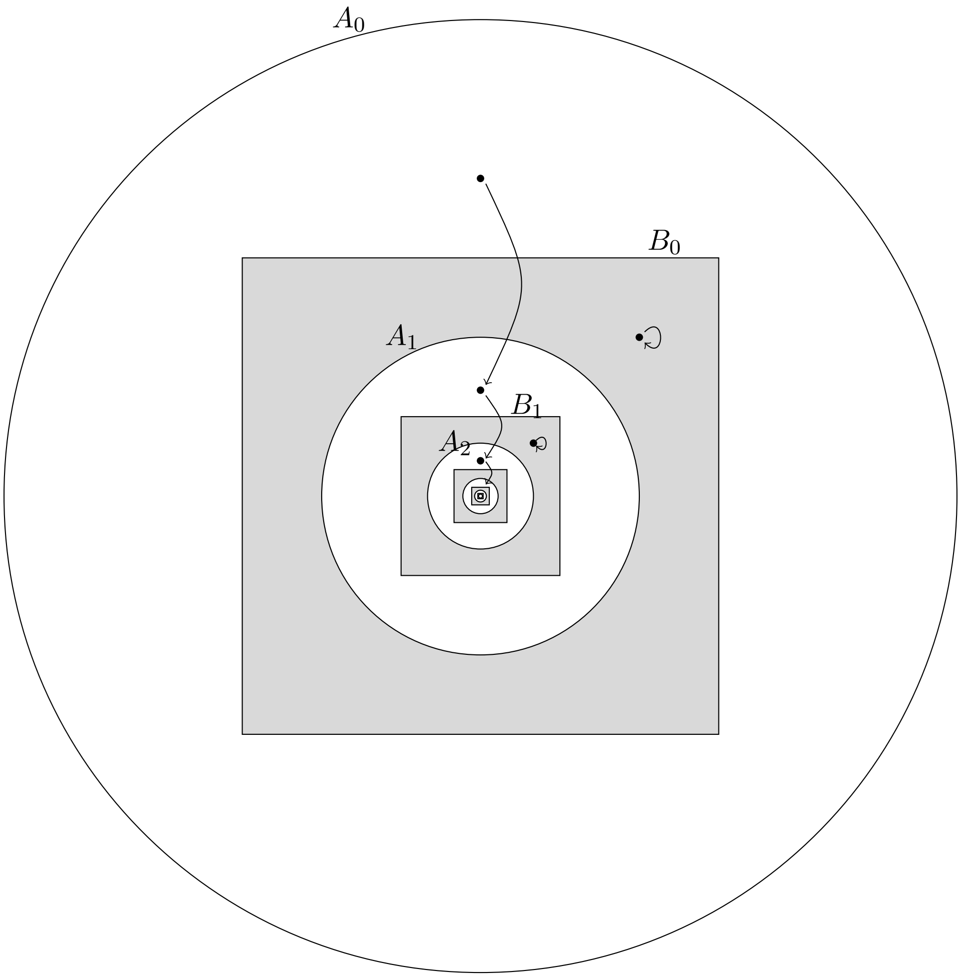

Figure 5.1: *Illustration of the Schröder–Bernstein Theorem.* The sets A_n are circles, and the sets B_n are rectangles. Furthermore the sets $A_n - B_n$ are shaded white, whereas the sets $B_n - A_{n+1}$ are shaded grey. The function f moves each point in a white region to the corresponding point in the next white region. It leaves points in the grey regions unmoved.

□

Right now, we do not yet know that there are uncountable sets. The next theorem, owing to Cantor, shows that there are:

Theorem 5.1.6 *For every set X, we have*

$$|X| < |\mathcal{P}(X)|$$

Proof: Clearly $X \to \mathcal{P}(X) : x \mapsto \{x\}$ defines an injection form X into $\mathcal{P}(X)$. Hence $|X| \leq |\mathcal{P}(X)|$.

Let $f : X \to \mathcal{P}(X)$ be an arbitrary function. We show that f cannot be a surjection: Indeed, define an element $Y \in \mathcal{P}(X)$ by

$$Y := \{x \in X : x \notin f(x)\}$$

If $Y \in \mathrm{ran}(f)$, then $Y = f(y)$ for some $y \in X$. Then

$$y \in Y \quad \Longleftrightarrow \quad y \in f(y) \quad \Longleftrightarrow \quad y \notin Y$$

— contradiction. Hence $Y \notin \mathrm{ran}(f)$, i.e. f is not surjective.

Since f was arbitrary, it follows that there is no surjection — and thus no bijection — from X onto $\mathcal{P}(X)$. Hence $|X| \neq |\mathcal{P}(X)|$.

$\dashv$

As a consequence, $\mathcal{P}(\omega)$ is an uncountable set.

5.2 Cardinal Arithmetic in ZF

We stated in the previous section that we do not (yet) have candidates for the cardinality $|X|$ of an infinite set X. We will now proceed *as if* we can assign a cardinal number $\kappa = |X|$ to a set X, and define arithmetic operations on these cardinals. Each statement about cardinal arithmetic can be translated into a statement about operations on sets: The notion of cardinal number is not necessary, but it is convenient. We use the letters $\kappa, \lambda, \mu, \nu$ to denote cardinals.

Recall that A^B is the set of all functions from B to A.

Definition 5.2.1 Suppose that $\kappa = |A|$, $\lambda = |B|$.

(a) $\kappa + \lambda := |A \cup B|$, where we require that $A \cap B = \varnothing$.

(b) $\kappa \cdot \lambda := |A \times B|$.

(c) $\kappa^\lambda := |A^B|$.

□

Remarks 5.2.2 (a) Note in the definition of cardinal addition that if A, B are not disjoint, one can always replace them with equivalent sets which are disjoint. Simply choose $x \neq y$, and define $A' := A \times \{x\}$, $B' := B \times \{y\}$. Then $|A| = |A'|$, $|B| = |B'|$, and $A' \cap B' = \varnothing$.

(b) Clearly $\kappa^0 = 1$ and , $1^\kappa = 1$. Furthermore $0^\kappa = 0$ if $\kappa > 0$. Observe, however, that there is a unique function $\varnothing \to \varnothing$, so that $0^0 = 1$.

Exercise 5.2.3 Check that the above definitions make sense, i.e. are independent of the choice of A, B. Thus check that

(a) If $A \cap B = \varnothing$, $\bar{A} \cap \bar{B} = \varnothing$, $|A| = |\bar{A}|$ and $|B| = |\bar{B}|$, then $|A \cup B| = |\bar{A} \cup \bar{B}|$.

(b) If $|A| = |\bar{A}|$ and $|B| = |\bar{B}|$, then $|A \times B| = |\bar{A} \times \bar{B}|$.

(c) If $|A| = |\bar{A}|$ and $|B| = |\bar{B}|$, then $|A^B| = |\bar{A}^{\bar{B}}|$.

□

The next proposition shows that cardinal arithmetic is better behaved than ordinal arithmetic.

Proposition 5.2.4 *(a)* $\kappa + \lambda = \lambda + \kappa$.

(b) $(\kappa + \lambda) + \mu = \kappa + (\lambda + \mu)$.

(c) $\kappa \cdot \lambda = \lambda \cdot \kappa$.

(d) $(\kappa \cdot \lambda) \cdot \mu = \kappa \cdot (\lambda \cdot \mu)$.

(e) $\kappa \cdot (\lambda + \mu) = \kappa \cdot \lambda + \kappa \cdot \mu$.

(f) $\kappa^{\lambda+\mu} = \kappa^\lambda \cdot \kappa^\mu$.

(g) $(\kappa \cdot \lambda)^\mu = \kappa^\mu \cdot \lambda^\mu$.

(h) If $\kappa \leq \lambda$*, then* $\kappa + \mu \leq \lambda + \mu$, $\kappa \cdot \mu \leq \lambda \cdot \mu$ *and* $\kappa^\mu \leq \lambda^\mu$.

(i) If $0 < \lambda \leq \mu$*, then* $\kappa^\lambda \leq \kappa^\mu$.

□

Exercise 5.2.5 Prove Proposition 5.2.4.
[E.g. to prove $\kappa^{\lambda+\mu} = \kappa^\lambda \cdot \kappa^\mu$, let $|A| = \kappa, |B| = \lambda, |C| = \mu$ with $B \cap C = \varnothing$, and show that there is a bijection from $A^{B \cup C}$ onto $A^B \times A^C$.]

□

Exercise 5.2.6 Recall from Exercise 5.1.3 that if X is an infinite countable set, then $|X| = |\omega|$. We thus denote the cardinality of an infinite countable set by the symbol ω. Show that $\omega + \omega = \omega \cdot \omega = \omega$.

□

The following theorem relates the cardinalities of power sets to cardinal arithmetic.

Theorem 5.2.7 *If* $|X| = \kappa$*, then* $|\mathcal{P}(X)| = 2^\kappa$*. Hence* $\kappa < 2^\kappa$ *for all cardinals* κ*.*

Proof: We have to show that there is a bijection from $\mathcal{P}(X)$ onto 2^X, where $2 := \{0,1\}$.

Observe that every subset $Y \subseteq X$ induces a unique *characteristic function* (or *indicator function*) $\chi_Y : X \to 2$ which, for each $x \in X$, indicates whether or not $x \in Y$. Thus:

$$\chi_Y(x) := \begin{cases} 1 & \text{if } x \in Y \\ 0 & \text{if } x \notin Y \end{cases} \qquad \text{for } x \in X$$

Conversely, every function $f : X \to 2$ indicates a unique subset $X_f := \{x \in X : f(x) = 1\}$. There is therefore a correspondence between subsets of X and functions $X \to 2$. To be precise, define

$$\Phi : \mathcal{P}(X) \to 2^X : Y \mapsto \chi_Y(\cdot) \qquad \Psi : 2^X \to \mathcal{P}(X) : f \mapsto X_f$$

Then it is easy to see that $\Phi \circ \Psi = \mathrm{id}_{2^X}$ and $\Psi \circ \Phi = \mathrm{id}_{\mathcal{P}(X)}$. Hence $\Psi = \Phi^{-1}$ and Φ is a bijection.

It follows by Theorem 5.1.6 that $\kappa < 2^\kappa$ for all cardinals κ.

$\dashv$

5.3 Alephs

We are now ready to define the notion of *cardinal number* for certain sets: The *alephs* are the cardinal numbers of *well–orderable sets*.

Recall that every well–ordering is order–isomorphic to a unique ordinal. Thus if X is well–orderable, then there is a bijection $\alpha \rightarrowtail\!\!\!\!\!\rightarrow X$ from some ordinal α to the set X. Conversely, if there is a bijection $f : \alpha \rightarrowtail\!\!\!\!\!\rightarrow X : \xi \mapsto x_\xi$, then f induces a well–ordering on X: Define $x_\xi < x_\eta$ if and only if $\xi < \eta$.

Thus a set is well–orderable if it and only if it can be placed into a bijective correspondence with some ordinal. Equivalently, a set X is well–orderable if and only if it can be *enumerated* by an ordinal, i.e. if and only if there is an $\alpha \in \mathbf{ON}$ such that $X = \{x_\xi : \xi < \alpha\}$.

The order–type of an enumeration need not be unique: At the beginning of this chapter, we showed how to enumerate the set of natural numbers by various ordinals: ω, $\omega + \omega$ and $\omega \times \omega + 1$. Since the size of the set of natural numbers is unaffected by how we order it, we have

$$|\omega| = |\omega + \omega| = |\omega \cdot \omega + 1|$$

Many other enumerations of the set of natural numbers are possible, but it is clear that the natural numbers cannot be enumerated by an ordinal which is $< \omega$: ω is the smallest ordinal which can enumerate the set of natural numbers. This leads to the following definition:

Definition 5.3.1 (a) An ordinal α is said to be a *cardinal number* (or just *cardinal*) if and only if for all ordinals β:

$$\beta < \alpha \quad \Longrightarrow \quad |\beta| < |\alpha|$$

(b) If a set X is a well–orderable set, then $|X|$ is the smallest ordinal which enumerates it, i.e.

$$|X| = \text{smallest ordinal } \alpha \text{ for which } |X| = |\alpha|$$

□

Thus α is a cardinal if and only if there is no bijection $f : \beta \to \alpha$ for any $\beta < \alpha$, i.e. if and only α cannot be enumerated by any $\beta < \alpha$. Moreover, if X is well—orderable, then $|X|$ is a cardinal.

An ordinal α which is a cardinal is also called an *initial ordinal*: It is the first ordinal to attain the size $|\alpha|$.

Clearly $0, 1, 2, \ldots, \omega$ are cardinal numbers. Furthermore, ω is the smallest infinite cardinal.

Remarks 5.3.2 (a) The set of cardinals is therefore a subclass of the set of ordinals. As such, it inherits the well–ordering on **ON**. Observe that if α, β are cardinals, then

$$\alpha < \beta \quad \text{in } \mathbf{ON} \qquad \Longleftrightarrow \qquad |\alpha| < |\beta|$$

By Definition 5.3.1, we see that $\alpha < \beta$ implies $|\alpha| < |\beta|$. Conversely, if α, β are cardinals and $|\alpha| < |\beta|$, then it cannot be the case that $\beta \leq \alpha$, and hence $\alpha < \beta$.

There is therefore no chance of confusion if we use the same symbol $<$ for the order relation on both the class of ordinals and the class of cardinals.

(b) We have defined arithmetic operations on both the class of ordinals and the class of cardinals. Here, however, care must be taken. The ordinal product $\omega \cdot \omega$ is not at all the same as the cardinal product $\omega \cdot \omega$. Indeed, for the ordinal product we have $\omega \cdot \omega > \omega$, yet $|\omega \cdot \omega| = |\omega|$. Thus the ordinal product $\omega \cdot \omega$ is an ordinal which is not a cardinal. The cardinal product is simply $\omega \cdot \omega = \omega$.

It is moreover not hard to see that for ordinal exponentiation, $\omega^\omega := \bigcup_{n \in \omega} \omega^n$ is countable, i.e. that for ordinal exponentiation ω^ω is a countable set. However, for cardinal exponentiation, we clearly have $|\omega|^{|\omega|} \geq 2^{|\omega|} > |\omega|$, i.e. for cardinal exponentiation, ω^ω is an uncountable set. The two interpretations of ω^ω do not even have the same cardinality!

Thus when doing arithmetic, you have to keep track of whether you are dealing with ordinals or cardinals. Only for the natural numbers do ordinal and cardinal arithmetic coincide.

□

Here is an easy result:

Proposition 5.3.3 *Every infinite cardinal is a limit ordinal.*

Proof: If $\alpha \geq \omega$ is infinite, then the map $f : \alpha + 1 \to \alpha$ defined by

$$f(\gamma) := \begin{cases} 0 & \text{if } \gamma = \alpha \\ \gamma + 1 & \text{if } \gamma < \omega \\ \gamma & \text{if } \omega \leq \gamma < \alpha \end{cases}$$

is a bijection, i.e $|\alpha + 1| = |\alpha|$. Thus $\alpha + 1$ is not a cardinal.

⊣

As we stated earlier, clearly $0, 1, 2, \ldots, \omega$ are cardinal numbers. Are there any others? We know that $|X| < |\mathcal{P}(X)|$. However, we generally do *not* know if $\mathcal{P}(X)$ is a well–orderable set, even if X is well–orderable. Thus though $|X|$ may be a well–defined cardinal number (in the sense of Definition 5.3.1), the object $|\mathcal{P}(X)|$ may not have any meaning in isolation.

Nevertheless, as we will now show, there are as many cardinals as there are ordinals. Since **ON** is a proper class, so is the class of cardinals.

Let X be any set (not necessarily well–orderable). Then $\mathcal{P}(X)$ is a set. Now if $Y \subseteq X$, then a well–ordering of Y is a subset of $Y \times Y$. Hence for each $Y \in \mathcal{P}(X)$, the class

$$\mathcal{W}_Y := \{\prec \in \mathcal{P}(Y \times Y) : \prec \text{ well–orders } Y\}$$

is a set, by Separation. It follows that $\mathcal{W} := \bigcup_{Y \in \mathcal{P}(X)} \mathcal{W}_Y$ is a set, the set of all possible well–orderings of subsets of X. We can define a class function F which assigns to each well–ordering $\prec \in \mathcal{W}$ the unique ordinal to which it is order–isomorphic. By the Axiom of Replacement, $F[\mathcal{W}]$ is a set. Thus $\bigcup F[\mathcal{W}] = \sup F[\mathcal{W}]$ is an ordinal, which may or may not belong to $F[\mathcal{W}]$. In any case, the successor ordinal of $\sup F[\mathcal{W}]$ cannot be a member of $F[\mathcal{W}]$. Thus $\{\alpha \in \mathbf{ON} : \alpha \notin F[\mathcal{W}]\}$ is a non–empty class of ordinals, and therefore has a least member — call it $h(X)$.

Now observe that, by definition, $\alpha \in F[\mathcal{W}]$ if and only if X has a subset which can be enumerated by α, i.e.

$$\alpha \in F[\mathcal{W}] \quad \Longleftrightarrow \quad \text{there is an injection from } \alpha \text{ to } X$$

Hence $h(X)$ is the least ordinal α for which there exists no injection from α into X.

The reasoning above shows that the ordinal $h(X)$ always exists, for any set X — well–orderable or not. Thus:

Definition 5.3.4 If X is a set, define the *Hartog's number* $h(X)$ of X as follows:

$$h(X) := \text{ least ordinal } \alpha \text{ such that there is no injection } \alpha \rightarrowtail X$$

□

Proposition 5.3.5 *For any set X, $h(X)$ is a cardinal number. Moreover, $h(X) \not\leq |X|$.*

Proof: If $\alpha < h(X)$, then there is an injection $\alpha \rightarrowtail X$. If it was the case that $|\alpha| = |h(X)|$, there would exist an injection $h(X) \rightarrowtail X$ — contradiction. Hence $|\alpha| < |h(X)|$ for all $\alpha < h(X)$.

Similarly, if $|h(X)| \leq X$, then there would exist an injection $h(X) \rightarrowtail X$ — contradiction, Hence $|h(X)| \not\leq |X|$.

$\dashv$

Observe that we are not saying that $h(X) > |X|$. That *would* be the case if the Law of Trichotomy holds, i.e. if (AC) holds. However, without (AC), it is possible that $|X|$, $h(X)$ are incomparable in size.

Proposition 5.3.6 *For every $\alpha \in$ **ON** there is a cardinal which is greater than α. Indeed, $h(\alpha)$ is the least cardinal $> \alpha$.*

Proof: Both $\alpha, h(\alpha)$ are ordinals. Thus either $h(\alpha) \leq \alpha$, or else $\alpha < h(\alpha)$. Since $|h(\alpha)| \not\leq |\alpha|$, we must have $\alpha < h(\alpha)$. Thus $h(\alpha)$ is a cardinal which is $> \alpha$.

If β is a cardinal and $\beta < h(\alpha)$, then there is an injection $\beta \rightarrowtail \alpha$, and hence $|\beta| \leq |\alpha|$. Because β is a cardinal, it then follows that $\beta \leq \alpha$. Hence $h(\alpha)$ is the least cardinal $> \alpha$.

⊣

If κ is a cardinal, then $h(\kappa)$ is the least cardinal $> \kappa$, i.e. the successor of κ in the class of cardinals. We denote it by

$$\kappa^+ := h(\kappa)$$

We already remarked that the order relation on the class of cardinals is a restriction of the order relation on the class of ordinals. Here is another result linking the two — The supremum of a set C of cardinals does not depend on whether you regard C as a set of ordinals or a set of cardinals:

Proposition 5.3.7 *If C is a set of cardinals, then $\sup C := \bigcup C$ is a cardinal.*

Proof: We know that $\alpha := \bigcup C$ is an ordinal, and that it is the least ordinal which is $\geq \kappa$ for every $\kappa \in C$. Now suppose that $\beta < \alpha$. By definition of α, there is $\kappa \in C$ such that $\beta < \kappa \leq \alpha$. Because κ is a cardinal, it follows that $|\beta| < |\kappa| \leq |\alpha|$. Hence $|\beta| < |\alpha|$ whenever $\beta < \alpha$.

⊣

We can now define an enumeration of the class of all infinite cardinals, the *alephs*[3]

$$\aleph_0, \aleph_1, \aleph_2, \ldots, \aleph_\omega, \aleph_{\omega+1}, \ldots, \aleph_{\aleph_1}, \ldots, \aleph_{\aleph_\omega}, \ldots$$

by transfinite recursion:

$$\begin{aligned} \aleph_0 &:= \omega_0 := \omega && \text{is the set of natural numbers} \\ \aleph_{\alpha+1} &:= \omega_{\alpha+1} := \aleph_\alpha^+ && \text{for successor ordinals } \alpha+1 \\ \aleph_\lambda &= \omega_\lambda := \sup\{\aleph_\alpha : \alpha < \lambda\} && \text{if } \lambda \text{ is a limit ordinal} \end{aligned}$$

We use $\aleph_\alpha$ and ω_α interchangeably. The $\aleph$–notation is typically used when we want to emphasize the *size* of a set, whereas the ω–notation when we are concerned also with the *order–type*.

Observe that ω_1 is the least uncountable ordinal, i.e. that $\alpha < \omega_1$ if and only if $|\alpha| \leq |\omega|$. Then ω_2 is the least ordinal whose cardinality is $> |\omega_1|$, etc.

Exercise 5.3.8 Show that the map $F : \mathbf{ON} \to \mathbf{ON} : \alpha \mapsto \omega_\alpha$ is *normal*, i.e. strcitly increasing and continuous. Deduce that $\omega_\alpha \geq \alpha$ for all $\alpha \in \mathbf{ON}$. Also conclude that there are arbitrarily large ordinals for which $\aleph_\alpha = \alpha$.

□

Next, we observe that $\{\aleph_\alpha : \alpha \in \mathbf{ON}\}$ enumerates *all* the infinite cardinals:

[3] Aleph $\aleph$ is the first letter of the Hebrew alphabet. The next two, Beth $\beth$, and Gimel $\gimel$ will also make an appearance in cardinal arithmetic.

Proposition 5.3.9 *If κ is an infinite cardinal, then there is a unique $\alpha \in \mathbf{ON}$ such that $\kappa = \aleph_\alpha$.*

Exercise 5.3.10 Prove Proposition 5.3.9.
[Hint: Show that if κ is a cardinal, then $\kappa < \aleph_{\kappa+1}$. Now show by induction that the following holds for all all α

$$\text{If } \kappa \text{ is a cardinal such that } \kappa < \aleph_\alpha, \text{ then } \kappa = \aleph_\beta \text{ for some ordinal } \beta < \alpha.$$

]

□

The alephs were introduced by Cantor. We emphasize once again that the alephs are the sizes (cardinalities) of *well–orderable* infinite sets. An infinite set X is well–orderable if and only if $|X| = \aleph_\alpha$ for some α. For Cantor, the restriction to well–orderable sets posed no problem: Already in 1883 he had proposed as a *Law of Thought* that every set is well–orderable. However, this proposal met with much opposition, and thus Cantor tried to find a proof. The first proof was, in fact, found by Zermelo, but his proof used a new principle for set formation — the Axiom of Choice. We shall have much more to say about this in the next chapter.

5.4 The Canonical Well–Ordering of ON × ON

Recall that, given a well–ordered set$(X, <)$, we defined in Lemma 4.1.2 the *canonical well–ordering* $(X \times X, \sqsubset)$. We can define the canonical well–ordering of the class $\mathbf{ON} \times \mathbf{ON}$ in the same manner: Given $\alpha, \beta, \alpha', \beta' \in \mathbf{ON}$, we put

$$\langle \alpha, \beta \rangle < \langle \alpha', \beta' \rangle \qquad \iff \qquad \begin{cases} & \max\{\alpha, \beta\} < \max\{\alpha', \beta'\} \\ \text{or} & \max\{\alpha, \beta\} = \max\{\alpha', \beta'\} \text{ and } \alpha < \alpha' \\ \text{or} & \max\{\alpha, \beta\} = \max\{\alpha', \beta'\},\ \alpha = \alpha',\ \text{and } \beta < \beta' \end{cases}$$

It is straightforward to verify that this does indeed define a well–ordering of $\mathbf{ON} \times \mathbf{ON}$, i.e. a total ordering with the property that every non–empty subclass of $\mathbf{ON} \times \mathbf{ON}$ has a least element.

Observe that if $\gamma \in \mathbf{ON}$, then the restriction $<\restriction (\gamma \times \gamma)$ of the canonical well–ordering on $\mathbf{ON} \times \mathbf{ON}$ to the set $\gamma \times \gamma$ is precisely the canonical well–ordering of $\gamma \times \gamma$.

Here are some observations that will be useful in the sequel:

- Observe that the initial segment $\{(\xi, \eta) : (\xi, \eta) < (\alpha, \beta)\}$ is a *set*, as it is a subset of $(\max\{\alpha, \beta\} + 1) \times \max\{\alpha, \beta\} + 1)$.

- Further observe that

  $$(\xi, \eta) < (0, \alpha) \quad \iff \quad \xi < \alpha \ \wedge\ \eta < \alpha \qquad \text{i.e.} \qquad \{(\xi, \eta) : (\xi, \eta) < (0, \alpha)\} = \alpha \times \alpha$$

  Thus each $\alpha \times \alpha$ is an *initial segment* of the canonical well–ordering of $\mathbf{ON} \times \mathbf{ON}$.

- Since every well–ordered set is order–isomorphic to a unique ordinal, we may define a map $\Gamma : \mathbf{ON} \times \mathbf{ON} \to \mathbf{ON}$ as follows: $\Gamma(\alpha, \beta)$ is the order–type of the initial segment induced by (α, β) in the canonical well–ordering of $\mathbf{ON} \times \mathbf{ON}$, i.e.

$$\Gamma(\alpha, \beta) := \text{order–type of } \{(\xi, \eta) : (\xi, \eta) < (\alpha, \beta)\}$$

- Γ is an order–preserving injection from $\mathbf{ON} \times \mathbf{ON}$ into $\mathbf{ON}$: No well–ordered set is order–isomorphic to an initial segment of itself.

- If $\beta < \alpha$, then $\Gamma(0, \beta) < \Gamma(0, \alpha)$: This is because $\Gamma(0, \alpha)$ is the order–type of of the initial segment $\alpha \times \alpha$ of $\mathbf{ON} \times \mathbf{ON}$. Since $\beta \times \beta$ is an initial segment of $\mathbf{ON} \times \mathbf{ON}$ and $\beta \times \beta \subseteq \alpha \times \alpha$, we see that $\beta \times \beta$ is an initial segment of $\alpha \times \alpha$.

- $\operatorname{ran}(\Gamma)$ is downwards–closed: If $\gamma \in \operatorname{ran}(\Gamma)$ and $\delta < \gamma$, then $\delta \in \operatorname{ran}(\Gamma)$. This is because some initial segment of $\mathbf{ON} \times \mathbf{ON}$ has order–type γ, and an initial segment of that initial segment must have order–type δ.

- Γ is surjective: The map $\Phi : \mathbf{ON} \to \mathbf{ON} : \alpha \to \Gamma(0, \alpha)$ is strictly increasing, and hence $\Phi(\alpha) \geq \alpha$ for all $\alpha \in \mathbf{ON}$, by Theorem 4.1.6. It follows that $\alpha \leq \Gamma(0, \alpha) \in \operatorname{ran}(\Gamma)$. Since $\operatorname{ran}(\Gamma)$ is downwards–closed, it follows that $\alpha \in \operatorname{ran}(\Gamma)$, for all $\alpha \in \mathbf{ON}$.

We reiterate: The map $\Gamma : \mathbf{ON} \times \mathbf{ON} \to \mathbf{ON}$ is an order–preserving bijection. We refer to Γ as the *canonical bijection* from $\mathbf{ON} \times \mathbf{ON}$ to $\mathbf{ON}$.

Exercise 5.4.1 Consider the map $\Psi : \mathbf{ON} \to \mathbf{ON} : \alpha \mapsto \Gamma(\alpha \times \alpha)$. Show that Ψ is a normal function. Conclude that $\Gamma(\alpha \times \alpha) \geq \alpha$ for all $\alpha \in \mathbf{ON}$. Further deduce that there are arbitrarily large ordinals for which $\Gamma(\alpha \times \alpha) = \alpha$.

□

With the previous exercise in mind, the next proposition should not be too surprising:

Proposition 5.4.2 $\Gamma(\omega_\alpha \times \omega_\alpha) = \omega_\alpha$ *for all* $\alpha \in \mathbf{ON}$.

Proof: Recall that $\omega_\alpha \times \omega_\alpha = \{(\xi, \eta) : (\xi, \eta) < (0, \omega_\alpha)\}$ is an initial segment of $\mathbf{ON} \times \mathbf{ON}$, and that $\Gamma(\omega_\alpha \times \omega_\alpha) = \Gamma(0, \omega_\alpha) \geq \omega_\alpha$. Suppose that α is the least ordinal such that $\Gamma(\omega_\alpha \times \omega_\alpha) \neq \omega_\alpha$, so that $\Gamma(\omega_\alpha \times \omega_\alpha) > \omega_\alpha$. Then there exists $(\xi, \eta) \in \omega_\alpha \times \omega_\alpha$ such that $\Gamma(\xi, \eta) = \omega_\alpha$. Since ω_α is a limit ordinal, we can find an ordinal δ so that $\xi, \eta < \delta < \omega_\alpha$. Then $\delta \times \delta$ is an initial segment of $\mathbf{ON} \times \mathbf{ON}$. Since $(\xi, \eta) \in \delta \times \delta$, we have $\omega_\alpha \subseteq \Gamma(\delta \times \delta)$. Since Γ is injective, it follows that $\omega_\alpha \leq |\Gamma(\delta \times \delta)| = |\delta \times \delta|$. Now $|\delta| = \omega_\beta$ for some $\beta < \alpha$, and by the minimality assumption $\Gamma(\omega_\beta \times \omega_\beta) = \omega_\beta$. Since Γ is injective, we see that $\omega_\beta = |\Gamma(\omega_\beta \times \omega_\beta)| = |\omega_\beta \times \omega_\beta|$. It follows that

$$\omega_\alpha \leq |\delta \times \delta| = |\omega_\beta \times \omega_\beta| = \omega_\beta$$

— contradiction, since $\omega_\beta < \omega_\alpha$. Hence there is no ordinal α such that $\Gamma(\omega_\alpha \times \omega_\alpha) \neq \omega_\alpha$.

$\dashv$

5.5 Arithmetic of Alephs

Cardinal addition and multiplication are trivial:

Theorem 5.5.1

$$\aleph_\alpha + \aleph_\beta = \aleph_\alpha \cdot \aleph_\beta = \max\{\aleph_\alpha, \aleph_\beta\}$$

Proof: Since $\Gamma(\omega_\alpha \times \omega_\alpha) = \omega_\alpha$ and Γ is injective, it follows that

$$\aleph_\alpha \cdot \aleph_\alpha = |\omega_\alpha \times \omega_\alpha| = |\Gamma(\omega_\alpha \times \omega_\alpha)| = |\omega_\alpha| = \aleph_\alpha$$

Thus

$$\aleph_\alpha \cdot \aleph_\alpha = \aleph_\alpha \qquad \text{for all } \alpha \in \mathbf{ON}$$

Now if $\gamma = \max\{\alpha, \beta\}$, then by Proposition 5.2.4 it follows that

$$\aleph_\gamma = \max\{\aleph_\alpha, \aleph_\beta\} \leq \aleph_\alpha + \aleph_\beta \leq \aleph_\gamma + \aleph_\gamma = 2 \cdot \aleph_\gamma \leq \aleph\gamma \cdot \aleph_\gamma = \aleph_\gamma$$

Similarly,

$$\aleph_\gamma = \max\{\aleph_\alpha, \aleph_\beta\} \leq \aleph_\alpha \cdot \aleph_\beta \leq \aleph\gamma \cdot \aleph_\gamma = \aleph_\gamma$$

⊣

Remarks 5.5.2 (a) It is important to note that the above theorem holds only for *alephs*. If the Axiom of Choice holds, then every infinite cardinal is an aleph, as we shall soon be able to show. However, the statement

$$\text{For all infinite } \kappa, \quad \kappa \cdot \kappa = \kappa$$

i.e. the statement that for all infinite sets A we have $|A \times A| = |A|$, is equivalent to (AC).

(b) Finite sums and products of alephs are now easy to calculate. Exponentiation is altogether more difficult. We know, for example, that $2^{\aleph_0} > \aleph_0$. We do not know if $2^{\aleph_0}$ is an aleph, i.e. if $2^{\aleph_0} = \aleph_\alpha$ for some $\alpha \in \mathbf{ON}$. With (AC), $2^{\aleph_0}$ will be an aleph. Even with (AC), however, we are not able to determine the value of α for which $2^{\aleph_0} = \aleph_\alpha$.

The axioms of ZFC place very few constraints on the possible values of α. α might be any of 1,2,3,.... It *cannot* be ω. It can be any of $\omega+1, \omega+2, \omega+3, \ldots$, or for that matter $\aleph_1, \aleph_2, \aleph_3, \ldots$. Again, we cannot have $\alpha = \aleph_\omega$.

The assumption that $\alpha = 1$ — i.e. that $2^{\aleph_0} = \aleph_1$ — is due to Cantor, and is known as the *Continuum Hypothesis.* Here, the word *continuum* refers to the set $\mathbb{R}$ of real numbers. The *Generalized Continuum Hypothesis* is the assertion that

$$2^{\aleph_\alpha} = \aleph_{\alpha+1} \qquad \text{for all } \alpha \in \mathbf{ON}$$

□

The next theorem shows what $2^{\aleph_0}$ has to do with the continuum.

Theorem 5.5.3

$$|\mathbb{R}| = 2^{\aleph_0}$$

□

Exercise 5.5.4 We prove the preceding theorem. Assume here that the basic properties of the sets $\mathbb{Q}, \mathbb{R}$ of rational and real numbers are known. (We have not yet established that $\mathbb{Q}, \mathbb{R}$ can be formalized inside ZFC. However, such a formalization must of course capture all these known properties.)

(a) Show that $|\mathbb{Q}| = \aleph_0$.

(b) Observe that if $r \in \mathbb{R}$, then $r = \sup\{q \in \mathbb{Q} : q < r\}$ — a known property. Conclude that $|\mathbb{R}| \leq 2^{\aleph_0}$.

(c) An element $f \in 2^\omega$ is a function $f : \omega \to \{0, 1\}$, which can be thought of as a sequence of 0's and 1's. With each such function is associated a real number whose decimal expansion consists of digits given by this sequence. Show that $2^{\aleph_0} \leq |\mathbb{R}|$.

(d) Finally, conclude that $|\mathbb{R}| = 2^{\aleph_0}$.

□

5.6 Project: Cardinal Arithmetic and The Axiom of Choice

Problem 1:
We prove a theorem of Tarski (1924):

Theorem: *In ZF, (AC) holds if and only if $\kappa^2 = \kappa$ for every infinite cardinal κ.*

1.1 First, explain why if (AC) holds, then $\kappa^2 = \kappa$ for every infinite cardinal.

1.2 Next, we show that if κ is an infinite cardinal and if $\aleph$ is an aleph, then

$$\kappa + \aleph = \kappa \cdot \aleph \qquad \text{implies} \qquad \kappa \leq \aleph \quad \text{or} \quad \aleph \geq \kappa$$

So suppose that $\kappa + \aleph = \kappa \cdot \aleph$. Let K be an infinite set such that $|K| = \kappa$, and let $(A, <)$ be a well–ordered set such that $|A| = \aleph$.

(i) Explain why there exist disjoint sets K_1 and A_1 such that $|K_1| = \kappa$, $|A_1| = \aleph$ and $K \times A = K_1 \cup A_1$.

(ii) Explain why exactly one of the following possibilities must be the case: Either

(1) $\exists k \in K \; \forall a \in A \; ((k, a) \in K_1)$, or

(2) $\forall k \in K \; \exists a \in A \; ((k, a) \in A_1)$.

(iii) Show that in case (1), we have $\aleph \leq \kappa$.

(iv) Suppose now that case (2) holds. For each $k \in K$, let a_k be the least element of A such that $(k, a_k) \in A_1$. Observe that $\{(k, a_k) : k \in K\} \subseteq A_1$. Conclude that $\kappa \leq \aleph$.

1.3 Now suppose that $\kappa^2 = \kappa$ for every infinite cardinal. Let X be any infinite set, with cardinality $\lambda = |X|$, and let $\aleph := h(X)$ be the Hartog's number of X. Observe that $\aleph$ is indeed an aleph, as the notation suggests.

(i) Show that $(\lambda + \aleph)^2 \geq \lambda \cdot \aleph$.

(ii) Conclude that $\lambda + \aleph = \lambda \cdot \aleph$.

(iii) Explain why we cannot have $\aleph \leq \lambda$. Conclude that $\lambda < \aleph$.

(iv) Deduce that X can be well–ordered. Hence explain why (AC) holds.

□

Problem 2:
We prove a theorem of Sierpiński (1947):

Theorem: *In ZF, (GCH) implies (AC).*

Recall that $\kappa < 2^\kappa$ for any cardinal κ. The Generalized Continuum Hypothesis (GCH) is the following assertion: If κ is an infinite cardinal, then there are no cardinals strictly between κ and 2^κ. To be precise

If κ, λ are infinite cardinals such that $\kappa \leq \lambda \leq 2^\kappa$, then either $\lambda = \kappa$ or $\lambda = 2^\kappa$.

2.1 We first prove, without assuming either (AC) or (GCH),

Lemma 1: *If κ is an infinite cardinal, and if $h(\kappa)$ is the Hartog's number of κ, then*

$$h(\kappa) \leq 2^{2^{2^\kappa}}$$

Let X be a set such that $|X| = \kappa$. Then $\subseteq$ is a partial order relation on $\mathcal{P}(X)$. Thus if $\mathcal{X} \subseteq \mathcal{P}(X)$, then $(\mathcal{X}, \subseteq)$ is a partial ordering. It may happen that $(\mathcal{X}, \subseteq)$ is a well–ordering, or it may not. Define

$$\mathbb{W} := \{\mathcal{X} \subseteq \mathcal{P}(X) : (\mathcal{X}, \subseteq) \text{ is a well–ordering}\}$$

(i) Show that $\mathbb{W} \subseteq \mathcal{P}(\mathcal{P}(X))$.

(ii) Define an equivalence relation $\approx$ on $\mathbb{W}$ by

$$\mathcal{X} \approx \mathcal{Y} \qquad \Longleftrightarrow \qquad (\mathcal{X}, \subseteq) \text{ is order–isomorphic to } (\mathcal{Y}, \subseteq)$$

and let

$$\mathbb{E} := \mathbb{W}/\approx$$

be the set of equivalence classes of $\approx$. Explain why $\mathbb{E} \subseteq \mathcal{P}(\mathcal{P}(\mathcal{P}(X)))$. Conclude that $|\mathbb{E}| \leq 2^{2^{2^\kappa}}$.

(iii) Each $E \in \mathbb{E}$ is a collection of all subsets of $\mathcal{P}(X)$ that have the the same order–type. Thus with each $E \in \mathbb{E}$ we can associate a unique ordinal α_E. Thus $\mathbb{E}$ can itself be regarded as a well–ordered set, with $E \prec F$ if and only if $\alpha_E < \alpha_F$. Conclude that $|\mathbb{E}|$ is an aleph.

(iv) Show that $|\mathbb{E}| \not\leq |X|$. Conclude that $|\mathbb{E}| \geq h(\kappa)$.
[Hint: Let η be the order–type of the well–ordered set $\mathbb{E}$. Suppose that $|\mathbb{E}| \leq |X|$. Then there is an injection $f : \eta \to X : \xi \mapsto x_\xi$. Define $X_\beta := \{x_\xi : \xi < \beta\}$, and $\mathcal{X} := \{X_\beta : \beta < \eta\}$. Observe that $(\mathcal{X}, \subseteq)$ has order–type η. The initial segment of $\mathbb{E}$ determined by the equivalence class of $\mathcal{X}$ then has order–type η. . .]

(v) Finally, conclude that $h(\kappa) \leq 2^{2^{2^\kappa}}$.

2.2 Now prove the following easy results in cardinal arithmetic, not assuming (AC) or (GCH):

(i) **Lemma 2:** *If $\kappa \geq \aleph_0$, then $2 \cdot \kappa = \kappa$.*
Remark: Saying $\kappa \geq \aleph_0$ is not the same as saying that κ is infinite, but that it is Dedekind infinite (i.e. has a countable subset).
[Hint: Show $1 + \kappa = \kappa$.]

(ii) **Lemma 3:** *If $\kappa \geq \aleph_0$, then $2^\kappa + \kappa = 2^\kappa$.*

(iii) **Lemma 4:** *If λ, κ are cardinals such that $2 \cdot \kappa = \kappa$ and $\lambda + \kappa \geq 2^\kappa$, then $\lambda \geq 2^\kappa$.*
[Hint: Choose disjoint sets K_1, K_2, L so that $|K_1| = |K_2| = \kappa$, $|L| = \lambda$. Justify the equations

$$|L \cup K_1| = \lambda + \kappa = 2^\kappa = 2^{\kappa+\kappa} = |\mathcal{P}(K_1 \cup K_2)|$$

Let $f : L \cup K_1 \rightarrowtail\!\!\!\!\rightarrow \mathcal{P}(K_1 \cup K_2)$ be a bijection. If $E \subseteq K_2$, define

$$E^* := E \cup \{x \in K_1 : x \notin f(x)\}$$

Show that the map $\mathcal{P}(K_2) \to \mathcal{P}(K_1 \cup K_2) : E \mapsto E^*$ is injective. Further show that each $E^* \notin f[K_1]$. Conclude that each $E^* = f(y)$ for some $y \in L$. Deduce that $|L| \geq |\mathcal{P}(K_2)| = 2^\kappa$.]

2.3 Henceforth, assume (GCH): Let $\kappa \geq \aleph_0$ be an infinite cardinal, and let $h(\kappa)$ be the Hartog's number of κ. Define a sequence $(\rho_n)_{n<\omega}$ of cardinals inductively as follows:

$$\rho_0 := \kappa, \qquad \rho_{n+1} = 2^{\rho_n}$$

Prove:

Lemma 5: *If $\kappa \geq \aleph_0$ and $h(\kappa) \leq \rho_{n+1}$, then either κ is an aleph, or $h(\kappa) \leq \rho_n$.*

(i) Observe that $2 \cdot \rho_n = \rho_n$, by Lemma 2.

(ii) Show that $\rho_n \leq h(\kappa) + \rho_n \leq 2^{\rho_n}$ for all n.

(iii) Conclude from (GCH) that we have twho possible cases: Either

(1) $h(\kappa) + \rho_n = \rho_n$, or else

(2) $h(\kappa) + \rho_n = 2^{\rho_n}$.

(iv) Show that in case (1), we have $h(\kappa) \leq \rho_n$.

(v) Show that in case (2), we have $h(\kappa) \geq 2^{\rho_n} \geq \kappa$ by Lemma 4. Conclude that κ is an aleph.

2.4 Finally, let X be any infinite set. Define $\kappa = 2^{\aleph_0 + |X|}$.

(i) Observe that Lemma 1 states that $h(\kappa) \leq \rho_3$. Also observe that $h(\kappa) \not\leq \rho_0$. Use Lemma 5 to conclude that κ is an aleph.

(ii) Deduce that X can be well–ordered. Hence explain why (AC) holds.

□

Chapter 6

The Axiom Of Choice

Recall:

Axiom of Choice: Every family of non–empty sets has a *choice function*: If $\mathcal{X}$ is a family of non–empty sets, then there is a function f on $\mathcal{X}$ which *chooses* from each $X \in \mathcal{X}$ an element $f(X) \in X$.

Exercise 6.0.1 Show by induction that every finite family $\mathcal{X}$ of non–empty sets has a choice function.
Thus (AC) is needed only for infinite $\mathcal{X}$.

□

6.1 Choice Functions, Partitions and Cartesian Products

Our first result shows that for (AC) to hold, it suffices that it holds for every every family $\mathcal{X}$ of *mutually disjoint* non–empty sets:

Proposition 6.1.1 *Let (AC)* be the following statement:*

> *(AC)*: If $\mathcal{X}$ is a partition (i.e. a family of non–empty pairwise disjoint sets), then there is a set C which contains precisely one element from each $X \in \mathcal{X}$ (i.e. $C \cap X$ is a singleton for every $X \in \mathcal{X}$).*

Then (AC) and (AC)* are equivalent in ZF.

Proof: Clearly (AC) implies (AC)*.

Conversely, suppose that $\mathcal{X}$ is a family of non–empty sets, not necessarily pairwise disjoint. For each $X \in \mathcal{X}$, let $\hat{X} := \{X\} \times X$. Then $\hat{\mathcal{X}} := \{\hat{X} : X \in \mathcal{X}\}$ is a family of pairwise disjoint sets. By (AC)*, there is a set C which contains precisely one element from each $\hat{X} \in \hat{\mathcal{X}}$. Then C is a set of ordered pairs, and thus a binary relation. By definition, if $(X, x) \in C$, then $x \in X$. Moreover, C is clearly a function: If $(X, x) \in C$ and $(X, y) \in C$ then $(X, x), (X, y) \in C \cap \hat{X}$, and hence $x = y$. It follows that C is a choice function on $\mathcal{X}$.

⊣

Proposition 6.1.2 *In ZF, the following are equivalent:*

(a) (AC)

(b) Every surjection has a right inverse.

Proof: Suppose that (AC) holds, and that $X \overset{f}{\twoheadrightarrow} Y$ is a surjection. For each $y \in Y$ define $X_y := f^{-1}[\{y\}]$. Then $\mathcal{X} := \{X_y : y \in Y\}$ is a family of non–empty sets which partitions X. By (AC), $\mathcal{X}$ has a choice function $h : \mathcal{X} \to \bigcup \mathcal{X}$ with the property that $h(X_y) \in X_y$ for all $y \in Y$. In particular, $f(h(X_y)) = y$. Define $g : Y \to X : y \mapsto h(X_y)$. Then $(f \circ g)(y) = f(h(X_y)) = y$.

Conversely, suppose that every surjection has a right inverse. Let $\mathcal{X}$ be a family of *disjoint* non–empty sets, and let $f : \bigcup \mathcal{X} \to \mathcal{X}$ be defined as follows:

$$\text{If } x \in \bigcup \mathcal{X}, \text{ then } f(x) \text{ is the unique } X \in \mathcal{X} \text{ so that } x \in X$$

Then f is a surjection, and thus has a right inverse $g : \mathcal{X} \to \bigcup \mathcal{X}$. Clearly $g(X) \in X$, i.e. g is a choice function for the partition $\mathcal{X}$. By Proposition 6.1.2, (AC) holds.

⊣

Recall that the cartesian product $A_0 \times A_1$ of two sets is defined to be the set of all ordered pairs $\langle a_0, a_1 \rangle$, where $a_0 \in A_0$ and $a_1 \in A_1$. We can then define products with more factors inductively:

$$\prod_{k=0}^{n+1} A_k := \prod_{k=0}^{n} A_k \times A_{n+1} \qquad \text{i.e.} \qquad A_0 \times A_1 \cdots \times A_n \times A_{n+1} := (A_0 \times A_1 \times \cdots \times A_n) \times A_{n+1}$$

However, this allows us to define products with only finitely many factors. How should we define a product $\prod_{i \in I} A_i$ with infinitely many factors, i.e. where the index set I has infinite cardinality? More generally, if $\mathcal{A}$ is a non–empty collection of sets, can we define $\prod \mathcal{A}$? Here the intention is that if $\mathcal{A} = \{A_i : i \in I\}$, then $\prod \mathcal{A} = \prod_{i \in I} A_i$ (just as $\bigcup \mathcal{A} = \bigcup_{i \in I} A_i$.)

Clearly we have $\prod_{i=0}^{1} A_i := \{\langle a_0, a_1 \rangle : a_0 \in A_0 \wedge a_1 \in A_1$. Thus we ought to have

$$\prod_{i \in I} A_i := \{\langle a_i : i \in I \rangle : \forall i \in I \; (a_i \in A_i)\}$$

The problem is that we have not yet defined ordered tuples $\langle a_i : i \in I \rangle$ of infinite length. However, we can clearly think of each such tuple as a function, i.e.

$$\langle a_i : i \in I \rangle \quad \text{is the function} \quad f : I \to \bigcup_{i \in I} A_i \quad \text{defined by} \quad f(i) = a_i$$

Thus the function f *chooses*, for each $i \in I$, an element $f(i) \in A_i$. Then

$$\prod_{i \in I} A_i := \{f : f \text{ is a function } I \xrightarrow{f} \bigcup_{i \in I} A_i \text{ such that } f(i) \in A_i \text{ for all } i \in I\}$$

i.e. the cartesian product $\prod_{i \in I} A_i$ is the set of all functions f which choose, for each $i \in I$, an element $f(i) \in A_i$.

There is an obvious natural bijection between the old definition of $A_0 \times A_1$ as a set of ordered pairs and the new definition as a set of functions $\{0, 1\} \to A_0 \cup A_1$: The ordered pair

$\langle a_0, a_1 \rangle \in A_0 \times A_1$ corresponds to the function $\{0,1\} \to A_0 \cup A_1 : 0 \mapsto a_0, 1 \mapsto a_1$. Henceforth we shall identify the two objects without further explanation. It will usually be clear from context which definition of product is meant.

Now if $\mathcal{A}$ is a non–empty non–indexed collection of sets, then we can index $\mathcal{A}$ by itself:

$$\mathcal{A} := \{X_A : A \in \mathcal{A}\} \qquad \text{where} \qquad X_A := A$$

Then $\prod \mathcal{A} = \prod_{A \in \mathcal{A}} X_A$ is the set of all functions $f : \mathcal{A} \to \bigcup \mathcal{A}$ with the property that $f(A) \in X_A$, i.e. with the property that $f(\mathcal{A}) \in A$. Thus if $f \in \prod \mathcal{A}$, then f is a function which chooses, for each $A \in \mathcal{A}$ an element $f(A) \in A$, i.e.:

Definition 6.1.3 $\prod \mathcal{A}$ is the set of all choice functions on $\mathcal{A}$.

□

The following proposition is now obvious:

Proposition 6.1.4 *The following statements are equivalent in ZF:*

(a) (AC)

(b) A product of a non–empty family of non–empty sets is non–empty: Whenever $\mathcal{A}$ is a non–empty family of non–empty sets, then $\prod \mathcal{A} \neq \varnothing$.

□

Proposition 6.1.4 is the reason that (AC) was also called the *multiplicative axiom* in the past.

6.2 Well–Ordering

Cantor proposed as an obvious *Law of Thought* the *Well–Ordering Principle*: Every set can be well–ordered.

The intuition behind this law was simple. Suppose that X is a set. Pick an element $x_0 \in X$. If any elements are left over, pick an element x_1 from $X - \{x_0\}$. If any are left over, pick an $x_2 \in X - \{x_0, x_1\}$. This defines a sequence by transfinite recursion: If x_ξ have been chosen for $\xi < \alpha$, then pick an element $x_\alpha \in X - \{x_\xi : \xi < \alpha\}$. Just keep doing this until you run out of elements. If this happens at stage β, then $X = \{x_\xi : \xi < \beta\}$. This enumeration of X induces well–ordering of X in an obvious manner.

The discovery of the paradoxes generated increased skepticism about the foundations of set theory, and this led Cantor to seek a proof of his Well–Ordering Principle. The first such proof was given by Zermelo in 1904 (and again in 1908). In doing so, Zermelo explicitly formulated the Axiom of Choice.

Proposition 6.2.1 *Let X be a set. In ZF the following are equivalent:*

(a) The set X can be well–ordered.

(b) The family $\mathcal{X} := \mathcal{P}(X) - \{\varnothing\}$ of non–empty subsets of X has a choice function.

Proof: (b) $\Rightarrow$ (a): Cantor's intuition can be implemented as follows: Suppose that F is a choice function for $\mathcal{X} := \mathcal{P}(X) - \{\varnothing\}$. Pick a set c such that $c \notin X$. We extend F to a function on $\mathbf{V}$ by defining $F(x) = c$ if $x \notin \mathcal{X}$. Define a function G on $\mathbf{ON}$ by transfinite recursion follows:

$$G(\alpha) = F(X - \operatorname{ran}(G \restriction \alpha))$$

Observe that this means that $G(\alpha) \in X - \{G(\xi) : \xi < \alpha\}$ as long as $X - \operatorname{ran}(G \restriction \alpha) \neq \varnothing$.

Now if $G(\alpha) \neq c$ for any $\alpha \in \mathbf{ON}$, then G is injective, and hence the function $G^{-1} : \operatorname{ran}(G) \to \mathbf{ON}$ exists, and is a bijection. Since $\operatorname{ran}(G) \subseteq X$, we see that $G^{-1}[X] = \mathbf{ON}$ is a set, by Replacement — contradiction. Hence there is $\alpha \in \mathbf{ON}$ such that $G(\alpha) = c$. If α_0 is the least such α, then $X - \operatorname{ran}(G \restriction \alpha_0) = \varnothing$, and hence $X = \{G(\xi) : \xi < \alpha_0\}$. Define a relation $\prec$ on X by

$$G(\xi) \prec G(\eta) \qquad \text{if and only if} \qquad \xi < \eta$$

Then $G : (\alpha_0, <) \to (X, \prec)$ is an order–isomorphism, so $\prec$ well–orders X.

(a) $\Rightarrow$ (b): If $(X, \prec)$ is a well–ordering, define F on $\mathcal{X} := \mathcal{P}(X) - \{\varnothing\}$ by:

$$F(Y) := \ \prec\text{–least element of } Y$$

To be precise F is the set

$$F = \{(Y, y) \in \mathcal{X} \times X : y \in Y \ \wedge \ \forall z \in Y \ (y \preceq z)\}$$

which exists by Separation.

Then F is clearly a choice function on $\mathcal{X}$.

$\dashv$

As an immediate corollary we have:

Theorem 6.2.2 *In ZF the following are equivalent:*

(a) (AC)

(b) Every set can be well–ordered.

Proof: (a) $\Rightarrow$ (b): Suppose that X is a set. By (AC), $\mathcal{P}(X) - \{\varnothing\}$ has a choice function. By proposition 6.2.1, X can be well–ordered.

(b) $\Rightarrow$ (a): Suppose that $\mathcal{X}$ is a family of non–empty sets. Let $X = \bigcup \mathcal{X}$. Then $\mathcal{X} \subseteq \mathcal{P}(X) - \{\varnothing\}$. Since X can be well-ordered, there is a choice function F for $\mathcal{P}(X) - \{\varnothing\}$. Then $F \restriction \mathcal{X}$ is a choice function for $\mathcal{X}$.

$\dashv$

Recall that the alephs $\aleph_\alpha$ were defined to be the cardinals of the well–orderable sets. (AC) implies that every set is well–orderable. In particular, for every infinite set X there is a unique $\alpha \in \mathbf{ON}$ such that $|X| = \aleph_\alpha$. Hence (AC) considerably simplifies cardinal arithmetic of all sets. We shall investigate cardinal arithmetic in ZFC in a later section.

6.3 Zorn's Lemma and Other Maximal Principles

> *We wish to show how, by introducing a certain axiom on sets of sets instead of the well–ordering theorem, one is enabled to make the proofs shorter and more algebraic.*
>
> — Max Zorn,
> *A Remark on Method in Transfinite Algebra*, 1935

Recall the following:

Definition 6.3.1 Let $(P, \leq)$ be a partially ordered set.

(i) If $X \subseteq P$, then an element $u \in P$ is an *upper bound* for X if $\forall x \in X\ (x \leq u)$.

(ii) An element $m \in P$ is called a *maximal element* of P if $\neg\exists x \in P\ (m < x)$.

(iii) A non–empty subset $C \subseteq P$ is called a *chain* in P if C is totally ordered by $\leq$, i.e. for all $x, y \in C$, either $x \leq y$ or $y \leq x$.

(iv) A chain C in P is said to be a *maximal chain* if it is not a proper subset of any other chain in P.

□

Theorem 6.3.2 *In ZF, the following are equivalent:*

(a) **(AC)**

(b) **Zorn's lemma** *(1935): If every chain in a non–empty partially ordered set* $(P, \leq)$ *has an upper bound, then* P *has a maximal element.*

(c) **Hausdorff's Maximal Principle** *(1909): Every chain in a non–empty partially ordered set can be extended to a maximal chain.*

Proof: (a) ⇒ (b): Let $(P, \leq)$ be a partially ordered set with the property that every chain in P has an upper bound, and let F be a choice function for the family $\mathcal{P}(P) - \{\varnothing\}$. Also choose some set $q \notin P$. For $A \subseteq P$, define $\upuparrows A := \{x \in P : \forall a \in A\ (a < x)\}$. By transfinite induction, define a map G on $\mathbf{ON}$ as follows: If $\alpha \in \mathbf{ON}$,

$$a_\alpha := G(\alpha) := \begin{cases} F(\upuparrows \operatorname{ran}(G \restriction \alpha) & \text{if } \upuparrows \operatorname{ran}(G \restriction \alpha) \neq \varnothing \\ q & \text{else} \end{cases}$$

To explain in simple English, suppose that $a_\xi := G(\xi)$ has been defined for $\xi < \alpha$. Then $a_\alpha := G(\alpha)$ is an element chosen (by F) from $\{x \in P : \forall \xi < \alpha\ (a_\xi < x)\}$, i.e. a_α is an element of P so that $a_\alpha > a_\xi$ for all $\xi < \alpha$, provided such an element exists. If such an element does not exists, then $a_\alpha := q$.

If $a_\alpha \neq q$ for all $\alpha \in \mathbf{ON}$, then $[\![a_\alpha : \alpha \in \mathbf{ON}]\!]$ is a sequence of distinct elements of P, i.e. $G : \mathbf{ON} \to P$ is an injection, which is impossible.[1] Let λ be the least ordinal such that

[1] By a similar argument as in the proof of Proposition 6.2.1. Here is another proof: If $h(P)$ is the Hartog's number of P, then $[\![a_\alpha : \alpha < h(P)]\!]$ would be an injection from $h(P)$ to P, contradicting the very definition of Hartog's number.

$a_\lambda = q$. Then $C := \{a_\xi : \xi < \lambda\}$ is a chain in P, and therefore has an upper bound u in P. We claim that u is a maximal element of P: For if not, then there is $x \in P$ such that $u < x$. Hence $\upuparrows \{a_\xi : \xi < \lambda\} \neq \varnothing$, so $a_\lambda \in P$, i.e. $q \in P$ — contradiction.

(b) ⇒ (c): Let C be a chain in a partially ordered set $(P, \leq)$. Let

$$\mathbb{P} := \{D \subseteq P : D \text{ is a chain in P, and } D \supseteq C\}$$

Then $(\mathbb{P}, \subseteq)$ is a partially ordered set, ordered by set inclusion. We show that every chain in $\mathbb{P}$ has an upper bound in $\mathbb{P}$. So suppose that $\mathbb{C} := \{C_i : i \in I\}$ is a chain in $\mathbb{P}$. Define $U := \bigcup \mathbb{C}$. Then

(i) U is a chain in $(P, \leq)$: For if $x, y \in U$, there are $i, j \in I$ such that $x \in C_i$ and $y \in C_j$. Since $\mathbb{C}$ is a chain in $(\mathbb{P}, \subseteq)$, we have either $C_i \subseteq C_j$ or $C_j \subseteq C_i$. Suppose without loss of generality that $C_i \subseteq C_j$. Then both $x, y \in C_j$. Since C_j is a chain in $(P, \leq)$, we have must have either $x \leq y$ or $y \leq x$.

(ii) $U \in \mathbb{P}$: Clearly $C \subseteq U$.

(iii) U is an upper bound of $\mathbb{C}$ in $\mathbb{P}$: If $C_i \in \mathbb{C}$, then $C_i \subseteq \bigcup \mathbb{C} = U$.

It follows from Zorn's Lemma that $\mathbb{P}$ has a maximal element C_M Then clearly C_M is a maximal chain in $(P, <)$ which extends C.

(c) ⇒ (a): We show that every family of non–empty sets has a choice function. Let $\mathcal{X}$ be a family of non–empty sets. Let $\mathbb{F}$ be the collection of all *partial choice functions*, i.e.

$$\mathbb{F} := \{f : f \text{ is a function, } \mathrm{dom}(f) \subseteq \mathcal{X}, \text{ and } f(X) \in X \text{ for all } X \in \mathrm{dom}(f)\}$$

Order $\mathbb{F}$ by inclusion, so that $(\mathbb{F}, \subseteq)$ is a partial ordering. By the Hausdorff Maximal Principle, there is a maximal chain $\mathbb{C}$ in $\mathbb{F}$. Let $F := \bigcup \mathbb{C}$. Then

(i) F is a function: For if $(X, x), (X, y) \in F$, then there exist $f, g \in \mathbb{C}$ so that $(X, x) \in f$ and $(X, y) \in g$. Since $\mathbb{C}$ is a chain in $(\mathbb{F}, \subseteq)$, either $f \subseteq g$ or $g \subseteq f$. Suppose without loss of generality that $f \subseteq g$. Then both $(X, x), (X, y) \in g$. Since g is a function, $x = y$, as required.

(ii) $F \in \mathbb{F}$: Indeed, if $(X, x) \in F$, there is $f \in \mathbb{C}$ so that $(X, x) \in f$. Observe that $F(X) = x = f(X)$. Since f is a partial choice function, $f(X) \in X$, and thus $F(X) \in X$.

(iii) $\mathrm{dom}(F) = \mathcal{X}$. For suppose there is $X \in \mathcal{X} - \mathrm{dom}(F)$. As each element of $\mathcal{X}$ is non–empty, one may choose $x \in X$. Then $G := F \cup \{(X, x)\} \in \mathbb{F}$ also. Moreover, $\mathbb{D} := \mathbb{C} \cup \{G\}$ is a chain, since $f \subseteq G$ for all $f \in \mathbb{F}$. Then $\mathbb{D}$ is a chain which properly extends $\mathbb{C}$ — contradiction, since $\mathbb{C}$ is a maximal chain.

Thus F is a choice function with domain $\mathcal{X}$.

⊣

Remarks 6.3.3 Various maximum principles were discovered in the years since Zermelo proposed the Axiom of Choice, by Hausdorff, Kuratowski, Tukey, Teichmüller, Zorn and others.

Zorn's Lemma is the (AC)–equivalent that has become the industry standard, a tool required by all mathematicians working with infinite sets. Although it goes by the moniker of "Lemma", Zorn actually proposed it as an axiom. Indeed, Zorn proposed the following maximum principle.

> (MP): Let $\mathcal{A}$ be a family of sets partially ordered by $\subseteq$ and closed under the taking of unions of $\subseteq$–chains. Then $\mathcal{A}$ has a maximum element.

We saw in Chapter 2 that every partially ordered set $(X, \leq)$ is order-isomorphic to a family of sets ordered by inclusion: Define $\mathcal{X} := \{\downarrow x : x \in X\}$, where $\downarrow x := \{y \in x : y \leq x\}$. Then $(X, \leq)$ and $(\mathcal{X}, \subseteq)$ are order–isomorphic. In particular, and chain in X corresponds to one in $\mathcal{X}$, any maximal element in X corresponds to one in $\mathcal{X}$, and vice versa. Thus the maximum principle (MP) is equivalent to Zorn's Lemma.

Observe that the proof of Zorn's lemma from (AC) required transfinite recursion. Zorn's lemma is a method by which one can avoid transfinite recursion — it is built in, as it were. This usually makes the technique more palatable to the working mathematician.

□

Here follows an example of the use of Zorn's lemma:

Theorem 6.3.4 *Every vector space has a basis.*

Let V be a vector space, and let $\mathbb{P}$ be the family of all linearly independent subsets of V. We claim that $(\mathbb{P}, \subseteq)$ satisfies the hypotheses of Zorn's Lemma. To see this let $\mathbb{C} := \{C_i : i \in I\}$ is a chain in $\mathbb{P}$. We must show that $\mathbb{C}$ has an upper bound. Define $U = \bigcup \mathbb{C}$. Then:

(i) $U \in \mathbb{P}$: For if $v_1, \dots, v_n \in U$, then there are $i_k \in I$ such that $v_k \in C_{i_k}$. Since $\mathbb{C}$ is a chain, the finite set $\{C_{i_1}, \dots, C_{i_n}\}$ has a maximum element, say C_{i_j}. Then each $v_1, \dots, v_n \in C_{i_j}$. Since C_{i_j} is a linearly independent subset of V, the vectors $v_1, \dots, v_n$ are linearly independent. Hence U is a linearly independent subset of V also.

(ii) U is an upper bound for $\mathbb{C}$ in $(\mathbb{P}, \subseteq)$ Each $C_i \in \mathbb{C}$ has $C_i \subseteq U$.

By Zorn's Lemma, $\mathbb{P}$ has a maximal element B, i.e. a maximal linearly independent subset. If $v \in V$ is such that $v \notin B$, then $B \cup \{v\}$ is not linearly independent. Hence there are $b_1, \dots, b_n \in B$ and scalars $\alpha_1, \dots, \alpha_n$ so that $\alpha_1 b_1 + \cdots + \alpha_n b_n + v = 0$. It is now easy to see that each $v \in V$ is a linear combination of members of B, so that B is a basis.

$\dashv$

We end with a short list of results that depend on (AC) for their validity.

- Every vector space has a basis. Moreover, any two bases have the same cardinality.
- *Tychonov's Theorem* in topology: The product $\prod_{i \in I} K_i$ of an arbitrary family of compact spaces is compact (in the product topology). Tychonov's Theorem is equivalent to (AC).

- *Nielsen–Schreier Theorem* in group theory: A subgroup of a free group is free.
- *Hahn–Banach Theorem* in functional analysis. This leads to various infinite–dimensional separating hyperplane theorems, which assert that any two disjoint closed convex sets can be separated by a hyperplane.
- Every field has a unique algebraic closure (up to isomorphism).
- *Compactness Theorem* in logic: If every finite subset of a set Σ of first–order sentences has a model, then Σ has a model.

6.4 Cardinal Arithmetic in ZFC

(AC) simplifies cardinal arithmetic considerably. It also extends its range to infinite sums and products.

Recall that, by definition, $|X| \leq |Y|$ if and only if there is an injection $X \rightarrowtail Y$. Every injection has a left inverse (without using (AC)) which is a surjection. Using (AC), every surjection has a right–inverse, which is an injection. Hence (using (AC)), $|X| \leq |Y|$ if and only if there is a surjection $Y \twoheadrightarrow X$.

Also recall that the alephs are the cardinals of well–orderable sets: If X is well–orderable, then there is an order isomorphism (and thus a bijection) from some ordinal to X. Then $|X|$ is defined to be the least ordinal for which there is a bijection onto X. Now, since the Axiom of Choice implies that every set can be well–ordered, *every* infinite cardinal $|X|$ is an aleph, i.e. for every infinite set X there is an $\alpha \in \mathbf{ON}$ such that $|X| = \aleph_\alpha$.

Thus the cardinals form a well–ordered class

$$0 < 1 < 2 < \ldots, < \aleph_0 < \aleph_1 < \ldots, < \aleph_\omega < \ldots$$

Every infinite cardinal appears on the list $[\![\aleph_\alpha : \alpha \in \mathbf{ON}]\!]$. i.e. the class of ordinals and the class of alephs are "order–isomorphic" as classes.

We defined the alephs inductively as follows

$$\begin{aligned} \aleph_0 &:= \omega \\ \aleph_{\alpha+1} &:= \aleph_\alpha^+ \quad \text{for successor ordinals } \alpha+1 \\ \aleph_\lambda &:= \sup\{\aleph_\alpha : \alpha < \lambda\} \quad \text{for limit ordinals } \lambda \end{aligned}$$

For obvious reasons, we call a cardinal of the form $\aleph_{\alpha+1}$ a *successor cardinal.* A cardinal of the form $\aleph_\lambda$ where λ is a limit ordinal, is called a *limit cardinal.* Every cardinal is either a limit cardinal or a successor cardinal.

For example $\aleph_0, \aleph_\omega, \aleph_{\omega_1}$ are limit cardinals, whereas $\aleph_1, \aleph_{\omega+1}, \aleph_{\omega_1+\omega\cdot3+5}$ are successor cardinals.

In particular, any two cardinals are *comparable* (this is the Law of Trichotomy): If X, Y are sets, then either $|X| \leq |Y|$ or $|Y| \leq |X|$. For if $|X| = \aleph_\alpha$ and $|Y| = \aleph_\beta$, then $\aleph_\alpha \leq \aleph_\beta$ if and only if $\alpha \leq \beta$.

Furthermore, as we showed earlier that $\aleph_\alpha + \aleph_\alpha = \aleph_\alpha \cdot \aleph_\alpha = \aleph_\alpha$, using the canonical well–ordering on $\mathbf{ON} \times \mathbf{ON}$, we must have $\kappa + \kappa = \kappa \cdot \kappa = \kappa$ for every infinite cardinal κ. It follows easily that

$$\kappa + \lambda = \kappa \cdot \lambda = \max\{\kappa, \lambda\}$$

for any two infinite cardinals κ, λ.

Here is an oft–used fact that relies on (AC):

Theorem 6.4.1 *(a) The union of a countable family of countable sets is countable.*

(b) If $\mathcal{X}$ is a family of sets, then

$$|\bigcup \mathcal{X}| \leq |\mathcal{X}| \cdot \sup\{|X| : X \in \mathcal{X}\}$$

Proof: (a) Let $\mathcal{X} := \{X_n : n < \omega\}$ be a countable collection of countable sets. For each $n < \omega$, there is a surjection $f_n : \omega \twoheadrightarrow X_n : m \mapsto x_{n,m}$ which lists the elements of X_n, i.e.

$$X_n := \{f_n(m) : m < \omega\} = \{x_{n,m} : m < \omega\}$$

Choose such an f_n for each $n < \omega$. We then have a surjection

$$f : \omega \times \omega \twoheadrightarrow \bigcup \mathcal{X} : (n, m) \mapsto x_{n,m}$$

Since we know $|\omega \times \omega| = \aleph_0 \cdot \aleph_0 = \aleph_0$, we see that $\bigcup \mathcal{X}$ is countable.

(b) Let $\kappa := |\mathcal{X}|$ and let $\lambda := \sup\{|X| : X \in \mathcal{X}\}$. Thus we can enumerate

$$\mathcal{X} := \{X_\alpha : \alpha < \kappa\}$$

Then, for each $\alpha < \kappa$, we can *choose* an enumeration

$$X_\alpha := \{x_{\alpha,\beta} : \beta < \lambda_\alpha\} \qquad \text{where } = \lambda_\alpha := |X_\alpha| \leq \lambda$$

Fix $x_0 \in \bigcup \mathcal{X}$, and define a map $p : \kappa \times \lambda \twoheadrightarrow \bigcup \mathcal{X}$ by

$$p(\alpha, \beta) \mapsto \begin{cases} x_{\alpha,\beta} & \text{if } \beta < \lambda_\alpha \\ x_0 & \text{else} \end{cases}$$

Then p is surjective, so $|\bigcup \mathcal{X}| \leq \kappa \cdot \lambda$.

$\dashv$

Proposition 6.4.2 *If $2 \leq \kappa \leq \lambda$ and λ is infinite, then*

$$\kappa^\lambda = 2^\lambda$$

Proof: $2^\lambda \leq \kappa^\lambda \leq (2^\kappa)^\lambda = 2^{\kappa \cdot \lambda} = 2^\lambda$, since $\kappa \cdot \lambda = \max\{\kappa, \lambda\} = \lambda$.

$\dashv$

It follows, for example, that $\aleph_0^{\aleph_0} = 2^{\aleph_0}$.

With (AC), we can define infinite sums and products of cardinals as follows:

Definition 6.4.3 (a) Let $\{\kappa_i : i \in I\}$ be an indexed set of cardinals. Let $\{X_i : i \in I\}$ be a family of pairwise disjoint sets such that $|X_i| = \kappa_i$. Define

$$\sum_{i \in I} \kappa_i := |\bigcup_{i \in I} X_i|$$

(b) Let $\{\kappa_i : i \in I\}$ be an indexed set of cardinals. Let $\{X_i : i \in I\}$ be a family of sets such that $|X_i| = \kappa_i$. Define

$$\prod_{i\in I} \kappa_i = |\prod_{i \in I} X_i|$$

where $\prod_{i\in I} X_i$ is the set of all choice functions, i.e. the set of all functions $f : I \to \bigcup_{i\in I} X_i$ with the property that $f(i) \in X_i$ for all $i \in I$.

(c) If κ, λ are cardinals, define

$$\kappa^{<\lambda} := \sup\{\kappa^\mu : \mu \text{ a cardinal}, \mu < \lambda\}$$

□

Remarks 6.4.4 Observe that without (AC), it does not follow that $\sum_{i\in I} \kappa_i$ is well–defined. One could have $|X_i| = |Y_i|$ for all $i \in I$, without having $|\bigcup_{i\in I} X_i| = |\bigcup_{i\in I} Y_I|$ (where $\{X_i : i \in I\}$ and $\{Y_i : i \in I\}$ are families of pairwise disjoint sets). The reason is that to construct a bijection which shows that $|\bigcup_{i\in I} X_i| = |\bigcup_{i\in I} Y_i|$, one first has to *choose* bijections $f_i : X_i \rightarrowtail\!\!\!\!\twoheadrightarrow Y_i$ that show $|X_i| = |Y_i|$.

Thus (AC) is necessary to make the definition of $\sum_{i\in I} \kappa_i$ independent of the sets X_i. A similar remark holds for $\prod_{i\in I} \kappa_i$.

□

Remarks 6.4.5 1. Observe that if $\kappa_i = \kappa$ for all $i \in I$, and if $|I| = \lambda$, then

$$\sum_{i\in I} \kappa_i = \kappa \cdot \lambda$$

i.e. sums and products behave as they ough to. For if X is a set of cardinality $|X| = \kappa$, and if we set $X_i := X \times \{i\}$ for all $i \in I$, then $\bigcup_{i\in I} X_i$ and $X \times I$ are the same set. Now by definition $\sum_{i\in I} \kappa_i = |\bigcup_{i\in I} X_i|$ and $\kappa \cdot \lambda = |X \times I|$.

2. Further observe that if $\kappa_i = \kappa$ for all $i \in I$, and if $|I| = \lambda$, then

$$\prod_{i\in I} \kappa_i = \kappa^\lambda$$

i.e. products and exponentials behave as they ought to: For if X is a set of cardinality $|X| = \kappa$, and if we set $X_i := X$ for all $i \in I$, then $\prod_{i\in I} X_i$ is the set of all functions from I to X, i.e. $\prod_{i\in I} X_i$ and X^I are the same set. Now by definition, $\prod_{i\in I} \kappa_i = |\prod_{i\in I} X_i|$ and $\kappa^\lambda = |X^I|$.

3. Next, observe that

$$\prod_{i\in I} \kappa_i^\lambda = (\prod_{i\in I} \kappa_i)^\lambda$$

For if $|X_i| = \kappa_i$, then there is an obvious bijection $\Psi : \prod_{i\in I} X_i^\lambda \to (\prod_{i\in I} X_i)^\lambda$ defined as follows. If $f \in \prod_{i\in I} X_i^\lambda$, then $f : I \to \bigcup_{i\in I} X_i^\lambda$ has $f_i := f(i) : \lambda \to X_i$ for each $i \in I$. Thus $f_i(\xi) \in X_i$ for each $i \in I$ and $\xi < \lambda$. Define $\hat{f} : \lambda \to \prod_{i\in I} X_i$ by $\hat{f}(\xi)(i) = f_i(\xi) = f(i)(\xi)$.

4. Similarly,
$$\prod_{i\in I}\kappa_i^\lambda = \kappa^{\sum_{i\in I}\lambda_i}$$

□

Proposition 6.4.6 *(a) Suppose that λ is an infinite cardinal. If $\kappa_i > 0$ for all $i < \lambda$, then*
$$\sum_{i<\lambda}\kappa_i = \lambda\cdot\sup_{i<\lambda}\kappa_i$$

(b) Suppose that λ is an infinite cardinal. If $[\![\kappa_i : i<\lambda]\!]$ is an increasing sequence of non–zero cardinals, then
$$\prod_{i<\lambda}\kappa_i = (\sup_{i<\lambda}\kappa_i)^\lambda$$

Proof: (a) Let $\sigma := \sum_{i<\lambda}\kappa_i$, and let $\mu := \sup\{\kappa_i : i<\lambda\}$. Clearly $\mu\leq\sigma$. Observe that
$$\sigma = \sum_{i<\lambda}\kappa_i \leq \sum_{i<\lambda}\mu = \lambda\cdot\mu \leq \lambda\cdot\sigma \qquad (\star)$$

Further note that, since each $\kappa_i > 0$,
$$\lambda = \sum_{i<\lambda}1 \leq \sum_{i<\lambda}\kappa_i = \sigma$$

Hence $\lambda\cdot\sigma = \sigma$, so by $(\star)$ it follows that $\sigma = \lambda\cdot\mu$.

(b) Let $\pi := \prod_{i<\lambda}\kappa_i$, and let $\mu := \sup\{\kappa_i : i<\lambda\}$. Observe that since the κ_i are non–zero, we have $\pi\geq\mu$. Now since $\kappa_i\leq\mu$ for all i, we have
$$\prod_{i<\lambda}\kappa_i \leq \prod_{i<\lambda}\mu = \mu^\lambda$$

To prove the reverse inequality, recall that $\sum_{i<\lambda}\lambda = \lambda\times\lambda = \lambda$. Thus we may partition λ into a family $\{A_j : j<\lambda\}$ of pairwise disjoint sets with cardinalities $|A_j| = \lambda$. Then
$$\pi = \prod_{i<\lambda}\kappa_i = \prod_{j<\lambda}(\prod_{i\in A_j}\kappa_i) = \prod_{j<\lambda}\mu = \mu^\lambda$$

(The fact that $\prod_{i<\lambda}\kappa_i = \prod_{j<\lambda}(\prod_{j\in A_j}\kappa_i)$ is easy to establish.)

⊣

If A is a set and κ a cardinal, define
$$[A]^\lambda := \{X\subseteq A : |X| = \lambda\} \qquad [A]^{<\lambda} := \{X\subseteq A : |X| < \lambda\}$$

Proposition 6.4.7 *(a) If $|A| = \kappa\geq\lambda$, then*
$$\left|[A]^\lambda\right| = \kappa^\lambda$$

(b) If $|A| = \kappa \geq \lambda$, and λ is an infinite cardinal, then

$$\left|[A]^{<\lambda}\right| = \kappa^{<\lambda}$$

Proof: (a) If $f \in A^\lambda$, then $f : \lambda \to A$, so $f \subseteq \lambda \times A$, and $|f| = \lambda$. Since $|\lambda \times A| = \lambda \cdot \kappa = \kappa$, we see that $\kappa^\lambda \leq |[\lambda \times A]^\lambda| = |[A]^\lambda|$. It follows that $\kappa^\lambda = |A^\lambda|$. On the other hand, for every $X \in [A]^\lambda$, there is at least one function $F_X : \lambda \to A$ whose range is X. Hence $[A]^\lambda| \leq |A^\lambda| = \kappa^\lambda$.

(b) We have

$$[A]^{<\lambda} = \bigcup_{\mu \text{ a cardinal } <\lambda} [A]^\mu$$

so by (a)

$$|[A]^\lambda| = \sum_{\mu \text{ a cardinal } <\lambda} \kappa^\mu = \lambda \cdot \sup\{\kappa^\mu : \mu \text{ a cardinal } < \lambda\} = \lambda \cdot \kappa^{<\lambda} = \kappa^{<\lambda}$$

⊣

Lastly, we state and prove:

Theorem 6.4.8 (König's Theorem) *If $\kappa_i < \lambda_i$ for all $i \in I$, then*

$$\sum_{i\in I} \kappa_i < \prod_{i \in I} \lambda_i$$

Proof: Suppose, on the contrary that $\prod_{i\in I} \lambda_i \leq \sum_{i\in I} \kappa_i$. Let X_i be sets such that $|X_i| = \lambda_i$. Since $|\prod_{i\in I} X_i| \leq \sum_{i\in I} \kappa_i$, we can find a partition $\{Y_i : i \in I\}$ of $\prod_{i\in I} X_i$ such that $|Y_i| \leq \kappa_i$. Here, the Y_i are pairwise disjoint, and $\bigcup_{i\in I} Y_i = \prod_{i\in I} X_i$.

For each $i \in I$, let P_i be the projection of Y_i onto the i^{th} coordinate, i.e.

$$P_i := \{f(i) : f \in Y_i\}$$

Observe that $P_i \subseteq X_i$. Now $|P_i| \leq |Y_i| = \kappa_i < \lambda_i = |X_i|$, so there is $x_i \in X_i - P_i$ for each $i \in I$. Define $h : I \to \bigcup_{i\in I} X_i : i \mapsto x_i$. Then certainly $h \in \prod_{i\in I} X_i$. Since $h(i) \notin P_i$ for any $i \in I$, we see that $h \notin Y_i$ for any $i \in I$. It follows that $h \in \prod_{i\in I} X_i - \bigcup_{i\in I} Y_i$ — contradiction.

⊣

Corollary 6.4.9 *For every cardinal κ, we have $\kappa < 2^\kappa$.*

Proof: $1 < 2$. Now

$$\kappa = \sum_{i<\kappa} 1 < \prod_{i<\kappa} 2 = 2^\kappa$$

⊣

Observe that we require the inequality $\kappa_i < \lambda_i$ to be strict in König's Theorem. For example, if $\kappa_n = \lambda_n = 2^{\aleph_0}$ for all $n < \aleph_0$, then

$$\sum_{n<\aleph_0} 2^{\aleph_0} = \aleph_0 \cdot 2^{\aleph_0} = 2^{\aleph_0}$$

and

$$\prod_{n<\aleph_0} 2^{\aleph_0} = (2^{\aleph_0})^{\aleph_0} = 2^{\aleph_0}$$

6.5 Length, Area, Volume and (AC)

"Length" is a non–negative real number associated with certain subsets of the real line $\mathbb{R}$, i.e. it is a function $\mathcal{L}(\cdot)$ which assigns to a subset $E \subseteq \mathbb{R}$ its length $\mathcal{L}(E)$. Intuitively, the length function should have certain properties:

$1_{\mathcal{L}}$. If $E \subseteq \mathbb{R}$ is bounded, then $0 \leq \mathcal{L}(E) < \infty$.

$2_{\mathcal{L}}$. $\mathcal{L}(\cdot)$ is *additive*: If E, F are disjoint bounded subsets of $\mathbb{R}$, then $\mathcal{L}(E \cup F) = \mathcal{L}(E) + \mathcal{L}(F)$.
More generally, $\mathcal{L}(\cdot)$ is *countably additive*: if $E_1, E_2, \ldots E_n, \ldots$ $(n \in \mathbb{N})$ is a sequence of mutually disjoint bounded subsets $\mathbb{R}$, then $\mathcal{L}(\bigcup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} \mathcal{L}(E_n)$.

$3_{\mathcal{L}}$. $\mathcal{L}(\cdot)$ is *translation invariant*: If a set F is obtained by shifting a set E, then $\mathcal{L}(F) = \mathcal{L}(E)$.

$4_{\mathcal{L}}$. If E is a interval, then $\mathcal{L}(E) =$ length of interval.

Using an argument due to Vitali in 1905, we show that it is impossible to assign a *length* to *every* bounded subset of $\mathbb{R}$, i.e. there is no function which satisfies each of the properties (1)-(4) of $\mathcal{L}(\cdot)$ above, and which is defined for every bounded subset of $\mathbb{R}$. Thus there are subsets of $\mathbb{R}$ which have no length. This does *not* mean that these sets have zero length; it means that there is no number which can be called their length, and which is consistent with (1)-(4).

Example 6.5.1 Define an equivalence relation $\sim$ on $\mathbb{R}$ by

$$x \sim y \qquad \Longleftrightarrow \qquad y - x \in \mathbb{Q}$$

Let $\{E_i : i \in I\}$ be an enumeration of the equivalence classes of $\sim$. Note that if $x \in \mathbb{R}$, then there exists $q \in \mathbb{Q}$ such that $0 \leq x + q \leq 1$. Now since $x \sim x + q$, we see that for every x there is $y \in [0,1]$ such that $x \sim y$. Thus $[0,1] \cap E_i \neq \emptyset$ for every $i \in I$.

Now pick[2] for each $i \in I$ one $x_i \in [0,1] \cap E_i$, and define a *Vitali set* H by $H := \{x_i :\in I\}$. Thus for each $y \in \mathbb{R}$ there is a unique $i \in I$ such that $y \sim x_i$.

For $q \in \mathbb{Q}$, define $H + q := \{x_i + q : i \in I\}$. First note that the $H + q$ are mutually disjoint: For if $y \in (H + q) \cap (H + q')$ for rational numbers q, q', then $y = x_i + q = x_j + q'$ for some $i, j \in I$, and thus $x_i = x_j + (q' - q)$, i.e. $x_i \sim x_j$. It follows that $x_i = x_j$, thus that $q = q'$, and thus that $H + q = H + q'$.

Next, we claim that for each $y \in \mathbb{R}$ there is a unique $q \in \mathbb{Q}$ such that $y \in H + q := \{x_i + q : i \in I\}$. Indeed, existence follows from the fact that there is an $i \in I$ such that $y \sim x_i$, so that $q := y - x_i$ has the property that $y \in H + q$. Uniqueness follows from the disjointness of the $H + q$.

Now let $\{q_n : n \in \mathbb{N}\}$ be an enumeration of $\mathbb{Q} \cap [-1,1]$. Note that if $x \in [0,1]$, there is a unique $i \in I$ such that $x - x_i \in \mathbb{Q} \cap [-1,1]$, so that $x \in \bigcup_{n \in \mathbb{N}} (H + q_n)$. Since $H \subseteq [0,1]$, we also have $\bigcup_{n \in \mathbb{N}} (H + q_n) \subseteq [-1,2]$. Thus:

$$[0,1] \subseteq \bigcup_{n \in \mathbb{N}} (H + q_n) \subseteq [-1,2]$$

[2] This requires the Axiom of Choice.

Now suppose that the Vitali set H has an length i.e. that $\mathcal{L}(H)$ exists. Each $H+q_n$ is a translation of H, and thus $\mathcal{L}(H+q_n)=\mathcal{L}(H)$ for all $n\in\mathbb{N}$. Now since $[0,1]\subseteq\bigcup_{n=1}^{\infty}(H+q_n)\subseteq[-1,2]$, it follows that

$$1\leq\mathcal{L}\left(\bigcup_{n=1}^{\infty}(H+q_n)\right)\leq 3$$

Furthermore, the sets $H+q_n$ are mutually disjoint, so

$$\mathcal{L}\left(\bigcup_{n=1}^{\infty}(H+q_n)\right)=\sum_{n=1}^{\infty}\mathcal{L}(H+q_n))=\sum_{n=1}^{\infty}\mathcal{L}(H)$$

The fact that $1\leq\sum_{n=1}^{\infty}\mathcal{L}(H)$ implies that $\mathcal{L}(H)>0$, whereas the fact that $\sum_{n=1}^{\infty}\mathcal{L}(H)\leq 3<\infty$ implies that $\mathcal{L}(H)\not>0$ — contradiction. Hence $H\subseteq\mathbb{R}$ is a bounded set to which a length cannot be assigned.

□

Just so, under (AC), some subsets of $\mathbb{R}^2$ fail to have an area, and some subsets of $\mathbb{R}^3$ fail to have a volume. The latter seems particularly counterintuitive: Given any subset of $\mathbb{R}^3$, to find its volume, just dump it in a full bucket of water, and measure how much water flows out.

It gets even worse, as will be shown in the next section.

6.6 Project: The Banach–Tarski Theorem

The paradoxical decomposition of the unit ball U in $\mathbb{R}^3$ due to Banach and Tarski (1924) is one reason why the Axiom of Choice may be regarded with some suspicion. It says that the unit ball U can be broken up into finitely many pieces (into five pieces, in fact), and that these pieces can be reassembled to form two new unit balls!

First, we introduce some terminology and notation so that we can phrase the Banach–Tarski Theorem.

- An isometry on $\mathbb{R}^3$ is a distance preserving map, i.e. a map $f:\mathbb{R}^3\to\mathbb{R}^3$ such that $||f(\mathbf{x})-f(\mathbf{y})||=||\mathbf{x}-\mathbf{y}||$ for all $\mathbf{x},\mathbf{y}\in\mathbb{R}^3$. Observe that any isometry is clearly a bijection.

  Translations and rotations (and compositions thereof) are isometries.

- Two sets $A,B\subseteq\mathbb{R}^3$ are *congruent* if and only if there is an isometry $f:\mathbb{R}^3\to\mathbb{R}^3$ so that $f[A]=B$. This means that A,B have the same shape — think of congruent triangles — and that f moves the figure A to the figure B.

  We write $A\equiv B$ if A,B are congruent sets.

- We say that a set A can be *reassembled* into a set B if there is a partition $(A_i)_{i<n}$ of A and a partition $(B_i)_{i<n}$ of B (into the same finite number of pieces) such that $A_i\equiv B_i$ for all $i<n$.

  In that case, we write $A\approx B$.

  The idea is that the set A can be broken up into pieces A_i, that each piece A_i can be moved by an isometry to B_i, and that the pieces B_i make up the set B. Thus A can be broken up into finitely many pieces, and the pieces can be reassembled to form B.

Theorem: (Banach–Tarski)
The unit ball U in $\mathbb{R}^3$ can be partitioned into two disjoint sets $U = U_1 \cup U_2$ such that

$$U_1 \approx U \qquad \text{and} \qquad U_2 \approx U$$

□

Problem 1.

(a) Show that $\approx$ is an equivalence relation on $\mathcal{P}(\mathbb{R}^3)$.

(b) Show that if $(A_i)_{i<n}$ is a partition of A and $(B_i)_{i<n}$ is a partition of B such that $A_i \approx B_i$ for all $i < n$, then $A \approx B$.

(c) Prove a "Schröder–Bernstein Theorem" for $\approx$: If $A^1 \subseteq B \subseteq A$ and $A^1 \approx A$, then $B \approx A$. [Hint: Let $(A_i)_{i<n}$ and $(A_i^1)_{i<n}$ be partitions of A and A^1, and let $f_i : A_i \to A_i^1$ be isometries so that $f_i[A_i] = A_i^1$. Define $f = \bigcup_{i<n} f_i$, so that f agrees with f_i on A_i. Define

$$A^0 = A, \qquad A^{n+1} = f[A^n], \qquad\qquad B^0 = B, \quad B^{n+1} = f[B^n]$$

Let $C := \bigcup_{n<\omega}(A^n - B^n)$. Observe that $f[C] \subseteq C$, and that $f[C] \approx C$. Further observe that $A = C \cup (A - C)$ and $B = f[C] \cup (A - C)$, and deduce that $A \approx B$.]

□

Our next aim is to partition the unit ball U into five pieces which can be reassembled to form two unit balls. To do that we will partition the unit sphere S into pieces having certain properties. (S is the boundary of U.) Let c denote the centre of the unit ball. Each $X \subseteq S$ determines a unique $\bar{X} \subseteq U$ as follows:

$\bar{X}$:= set of all points in $U - \{c\}$ whose projection along a ray
from c to the surface S lies in X.

Clearly, if $X, Y \subseteq S$ are such that $X \approx Y$, then also $\bar{X} \approx \bar{Y}$.

Consider now two axes d_0, d_1 in $\mathbb{R}^3$ centred at c. For definiteness, assume that d_0 is the Z–axis, and that d_1 lies in the XZ–plane at an angle θ to the Z–axis. Consider two rotations:

$$\psi = \text{ Rotation by } \pi/3 \text{ about } d_0 \qquad \varphi = \text{ Rotation by } \pi \text{ about } d_1$$

Observe that ψ, φ are isometries which map S to S.

Let I be the identity. Note that $\psi^3 = I$ and $\varphi^2 = I$, so that $\varphi^{-1} = \varphi, \psi^{-1} = \psi^2$. Consider the group G of all rotations generated by ψ, φ. Take, for example, the element $\alpha := \varphi \circ \varphi \circ \psi \circ I \circ \psi \circ \varphi$. Using the facts that $\psi^3 = I$ and $\varphi^2 = I$, we can reduce this to $\alpha = \psi^{-1} \circ \varphi$. In this way, every element of G has an irreducible representation of the form

$$\varphi \circ \psi^{\pm 1} \circ \varphi \circ \psi^{\pm 1} \circ \dots \qquad \text{or} \qquad \psi^{\pm 1} \circ \varphi \circ \psi^{\pm 1} \circ \varphi \circ \dots$$

Each such an irreducible representation has a *length*:

$$l(I) = 0, \quad l(\varphi) = l(\psi^{\pm 1}) = 1, \qquad l(\varphi \circ \psi^{\pm 1}) = l(\psi^{\pm 1} \circ \varphi) = 2, \qquad l(\varphi \circ \psi^{\pm 1} \circ \varphi) = 3 \dots$$

Now the group G depends on the angle θ between d_0 and d_1. We want to show that there is θ so that every element of G has a *unique* irreducible representation. Equivalently, we want to show that we can choose θ so that no non–trivial irreducible representation yields the identity I.

Problem 2.

(a) Show that each element of $\alpha \in G$ can be represented by an orthogonal matrix (rotation matrix) whose entries are bivariate polynomials in $\cos\theta$ and $\sin\theta$ (i.e. for each entry α_{ij} of the matrix α there is a polynomial $p_{ij}(x, y)$ so that $\alpha_{ij} = p_{ij}(\cos\theta, \sin\theta)$).
[Hint: A rotation by π about d_1 can be effected by rotating the axis d_1 by θ onto the Z–axis, then rotating by π, and then rotating the axis by $-\theta$ back to its original position.]

(b) Show that an irreducible representation $\alpha \in G$ of length ≥ 1 yields the identity element if and only if $z := e^{i\theta}$ is a solution to a system of polynomials.
[Hint: $\cos\theta = \frac{z+\frac{1}{z}}{2}$.]

(c) Deduce that for each $\alpha \in G$ there are at most finitely many θ so that α yields the identity element. Conclude that there exists a θ so that no irreducible representation of length ≥ 1 yields the identity element.

□

Remarks: We've tried to avoid group theory in the above. What we were actually doing, should you know some group theory, is to show that the free product $\mathbb{Z}_2 * \mathbb{Z}_3$ of the groups $\langle\varphi\rangle = \mathbb{Z}_2$ and $\langle\psi\rangle = \mathbb{Z}_3$ is embeddable into the group $SO(\mathbb{R}^3)$. If G_θ is the group of rotations generated by φ, ψ when the angle between d_0, d_1 is θ, then there is an obvious homomorphism $\pi_\theta : \mathbb{Z}_2 * \mathbb{Z}_3 \to G_\theta$. We showed that there exists θ such that π_θ is an isomorphism.

□

Henceforth, fix a θ so that no irreducible representation of length ≥ 1 yields the identity of the corresponding group G. Then every element of G has a unique irreducible representation. We define a partition $G = A \cup B \cup C$ by induction on the length of irreducible representations. The elements of length ≤ 1 are classified as follows:

$$I \in A, \qquad \varphi, \psi \in B \qquad \psi^{-1} \in C$$

Now classify the elements of length ≥ 2 according to the following rules:

- If α starts with $\psi^{\pm 1}$, then

  (1) $\alpha \in A \implies \varphi \circ \alpha \in B$.

  (2) $\alpha \in B \cup C \implies \varphi \circ \alpha \in A$.

  Observe that if α starts with φ, i.e. if $\alpha = \varphi \circ \beta$, then $\varphi \circ \alpha$ has irreducible representation β, since $\varphi \circ \alpha = \varphi \circ \varphi \circ \beta = \beta$. Since the length of β is $\leq$ the length of α, $\varphi \circ \alpha = \beta$ has already been classified.

- If α starts with φ, then

  (3) $\alpha \in A \implies \psi \circ \alpha \in B$ and $\psi^{-1} \circ \alpha \in C$.

  (4) $\alpha \in B \implies \psi \circ \alpha \in C$ and $\psi^{-1} \circ \alpha \in A$.

  (5) $\alpha \in C \implies \psi \circ \alpha \in A$ and $\psi^{-1} \circ \alpha \in B$.

  If α starts with $\psi^{\pm 1}$, then $\psi^{\pm 1} \circ \alpha$ has already been classified.

Problem 3: Show that

$$\varphi[A] = B \cup C \qquad \psi[A] = B \qquad \psi^{-1}[A] = C$$

[Hint: To show $\varphi[A] \subseteq B \cup C$, let $\alpha \in A$. If α starts with φ, then $\alpha = \varphi \circ \beta$ for some $\beta \in B \cup C$, by rule (2). Hence $\varphi \circ \alpha = \beta \in B \cup C$. Else, if α starts with $\psi^{\pm 1}$, then by rule (1), $\varphi \circ \alpha \in B \subseteq B \cup C$. Hence $\varphi \circ \alpha \in B \cup C$ for all $\alpha \in A$.

Conversely, to show $B \cup C \subseteq \varphi[A]$, suppose that $\alpha \in B \cup C$. If α starts with φ, then $\alpha = \varphi \circ \beta$ for some $\beta \in A$ (and thus $\varphi \circ \beta \in B$), so $\alpha \in \varphi[A]$. Else if α starts with $\psi^{\pm 1}$, then $\varphi \circ \alpha \in A$, by (2), so $\alpha = \varphi \circ (\varphi \circ \alpha) \in \varphi[A]$. Hence $\alpha \in \varphi[A]$ for all $\alpha \in B \cup C$.
The other equations can be proved similarly.]

□

We now use the partition $G = A \cup B \cup C$ to construct a partition $S = X \cup Y \cup Z \cup Q$ of the unit sphere. This, in turn, will induce a partition of the unit ball B.

Problem 4: We show that there is a partition $S = X \cup Y \cup Z \cup Q$ of the unit sphere S, where

$$|Q| = \aleph_0 \qquad \text{and} \qquad X \equiv Y \equiv Z \equiv Y \cup Z$$

(a) Observe that $|G| = \aleph_0$. Also note that every rotation of S other than the identity has exactly two fixed points. Let Q be the set of all $x \in S$ so that x is a fixed point of some $\alpha \in G - \{I\}$. Conclude that $|Q| = |\aleph_0|$.

(b) For $x \in S - Q$, let

$$P_x := \{\alpha(x) : \alpha \in G\}$$

Define a binary relation R on $S - Q$ by $xRy \longleftrightarrow x \in P_y$. Show that R is an equivalence relation.

(c) The Axiom of Choice implies that there is a set M which contains exactly one element of each equivalence class of R. Define

$$X := A \cdot M := \{\alpha(x) : x \in M, \alpha \in A\}, \qquad Y := B \cdot M, \qquad Z = C \cdot M$$

where $G = A \cup B \cup C$ is the partition of G constructed earlier. Show that

$$\varphi[X] = Y \cup Z, \qquad \psi[X] = Y, \qquad \psi^{-1}[X] = Z$$

(d) Conclude that $X \equiv Y \equiv Z \equiv Y \cup Z$.

□

Now recall that if $X \subseteq S$, then $\bar{X}$ is the set of all points in $U - \{c\}$ whose projection along a ray from the centre c onto the surface S lies in X. The partition $S = X \cup Y \cup Z \cup Q$ of S therefore induces a partition

$$U = \bar{X} \cup \bar{Y} \cup \bar{Z} \cup \bar{Q} \cup \{c\}$$

of the unit ball. Moreover,clearly

$$\bar{X} \approx \bar{Y} \approx \bar{Z} \approx \bar{Y} \cup \bar{Z}$$

Problem 5:

(a) Show that $\bar{Y} \cup \bar{Z} \approx \bar{X} \cup (\bar{Y} \cup \bar{Z})$. Deduce that

$$\bar{X} \approx \bar{Y} \approx \bar{Z} \approx \bar{Y} \cup \bar{Z} \approx \bar{X} \cup \bar{Y} \cup \bar{Z}$$

(b) Show that $U \approx \bar{X} \cup \bar{Q} \cup \{c\}$.

(c) Show that there is a rotation $\pi \notin G$ such that $\pi[Q] \cap Q = \varnothing$. Conclude that $\pi[Q] \subseteq X \cup Y \cup Z$. Using the fact that $\bar{Z} \approx \bar{X} \cup \bar{Y} \cup \bar{Z}$, deduce that there is $T \subseteq Z$ such that $Q \approx T$.
[Hint: Choose an axis of rotation whose poles are not members of Q. For each angle $\delta \in [0, 2\pi)$, let π_δ denote a rotation by δ about this axis. For $q \in Q$, define
$\Delta_q := \{\delta \in [0, 2\pi) : \pi_\delta(q) \in Q\}$. Then $\Delta := \bigcup_{q \in Q} \Delta_q$ is countable. Choose $\delta \in [0, 2\pi) - \Delta$, and let $\pi := \pi_\delta$.]

(d) Let $p \in \bar{Z} - \bar{T}$. Observe that

$$U \approx \bar{X} \cup \bar{Q} \cup \{c\} \approx \bar{Y} \cup \bar{T} \cup \{p\} \subseteq \bar{Y} \cup \bar{Z} \subseteq U$$

Conclude that $\bar{Y} \cup \bar{Z} \approx U$.

(e) Finally note that $U = U_1 \cup U_2$, where $U_1 := \bar{X} \cup \bar{Q} \cup \{c\}$ and $U_2 := \bar{Y} \cup \bar{Z}$. This concludes the proof of the Banach–Tarski Theorem.

□